

528,344

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



Rec'd PCT/PTO

13 MAR 2005



(43) Date de la publication internationale
1 avril 2004 (01.04.2004)

PCT

(10) Numéro de publication internationale
WO 2004/027068 A2

(51) Classification internationale des brevets⁷ : C12N 15/67

(74) Mandataire : AUDIER, Philippe; c/o Brevatome, 3, rue
du Docteur Lancereaux, F-75008 Paris (FR).

(21) Numéro de la demande internationale :

PCT/FR2003/002763

(81) États désignés (*national*) : AE, AG, AL, AM, AT, AU, AZ,
BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ,
DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC,
SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(22) Date de dépôt international :

19 septembre 2003 (19.09.2003)

(25) Langue de dépôt :

français

(26) Langue de publication :

français

(30) Données relatives à la priorité :

02/11676 20 septembre 2002 (20.09.2002) FR

(84) États désignés (*régional*) : brevet ARIPO (GH, GM, KE,
LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), brevet
eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet
européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK,
TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

(71) Déposant (*pour tous les États désignés sauf US*) :
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE [FR/FR]; 3, rue Michel Ange, F-75794 Paris
Cedex 16 (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (*pour US seulement*) : FALSON,
Pierre [FR/FR]; 5, avenue Valioud, F-69110 Sainte Foy
les Lyon (FR). PENIN, François [FR/FR]; 20, avenue des
Platanes, F-69150 Decines Charpieu (FR). MONTIGNY,
Cédric [FR/FR]; Résidence La Cilof, Bât. K., Route de
Chateaufort, F-91190 Gif sur Yvette (FR).

Publiée :

— sans rapport de recherche internationale, sera republiée
dès réception de ce rapport

En ce qui concerne les codes à deux lettres et autres abrévia-
tions, se référer aux "Notes explicatives relatives aux codes et
abréviations" figurant au début de chaque numéro ordinaire de
la Gazette du PCT.

(54) Title: SYSTEMS FOR EXPRESSING TOXIC PROTEINS, VECTORS AND METHOD OF PRODUCING TOXIC PRO-
TEINS

(54) Titre : SYSTEMES D'EXPRESSION DE PROTEINES TOXIQUES, VECTEURS ET PROCEDE DE FABRICATION DE
PROTEINES TOXIQUES

(57) Abstract: The invention relates to a system for expressing toxic proteins, an expression vector comprising the aforementioned system, a prokaryote cell transformed by the system and a method of synthesising a toxic protein using said expression system. Said expression system is characterised in that it comprises the following successive sequences in direction 5' 3', namely a nucleotide sequence coding for dipeptide Asp-Pro and a nucleotide sequence coding for a toxic protein. In a particularly preferable mode, the expression system also comprises, upstream of the Asp-Pro sequence, a nucleotide sequence coding for a soluble protein. The inventive expression system can be used to construct an expression vector which can be used to transform a prokaryote cell such as *E. coli*, e.g. in a toxic protein synthesis method.

(57) Abrégé : La présente invention se rapporte à un système d'expression de protéines toxiques, à un vecteur d'expression comprenant ce système, à une cellule procaryote transformée par ce système, ainsi qu'à un procédé de synthèse d'une protéine toxique utilisant ce système d'expression. Le système d'expression de l'invention se caractérise en ce qu'il comprend successivement, dans le sens 5'-3', une séquence nucléotidique codant pour le dipeptide Asp-Pro et une séquence nucléotidique codant pour une protéine toxique. Selon un mode particulièrement préféré de l'invention, le système d'expression comprend en outre, en amont de la séquence Asp-Pro, une séquence nucléotidique codant pour une protéine soluble. Le système d'expression de l'invention permet de construire un vecteur d'expression utile pour la transformation d'une cellule procaryote telle que *E. coli*, par exemple dans un procédé de synthèse de la protéine toxique.

WO 2004/027068 A2

**SYSTEMES D'EXPRESSION DE PROTEINES TOXIQUES, VECTEURS
ET PROCEDE DE FABRICATION DE PROTEINES TOXIQUES**

DESCRIPTION

5 DOMAINE TECHNIQUE

La présente invention se rapporte à des systèmes d'expression de protéines toxiques, à des vecteurs d'expression comprenant un de ces systèmes, à des cellules procaryotes transformées par ces systèmes, ainsi qu'à un procédé de synthèse d'une protéine toxique utilisant ces systèmes d'expression.

Elle permet par exemple la surproduction dans une cellule procaryote, par exemple *Escherichia coli* (*E.coli*), de protéines ou peptides hydrophobes toxiques, par exemple la surproduction de domaines transmembranaires de protéines d'enveloppe de virus.

Elle trouve de nombreuses applications notamment dans la recherche concernant les mécanismes d'infections virales et dans recherche et la mise au point de nouveaux principes actifs pour lutter contre les infections virales.

Dans la description qui suit, les références entre crochets [] renvoient à la liste de références annexée.

25

Etat de la technique

La détermination de la structure tridimensionnelle (3D) est une étape décisive dans la compréhension structurale et fonctionnelle des protéines.

30

De très grands efforts et moyens ont été, et sont, mis en oeuvre pour parvenir à ce but, et se sont amplifiés avec l'accumulation des données apportées par les programmes de séquençage des génomes [1].

5 Les deux principales techniques permettant l'établissement de ces structures des protéines sont la diffraction des rayons X, réalisée à partir de protéines cristallisées, et la résonance magnétique nucléaire (RMN) effectuée à partir de protéines en
10 solution. La RMN, bien adaptée à l'étude de protéines de masse moléculaire inférieure à 20 kDa, nécessite cependant, comme la diffraction des RX, la production de grandes quantités de matériel. Elle implique en outre dans la plupart des cas de préparer du matériel
15 enrichi en ^{15}N et/ou ^{13}C .

Dans ce contexte, la bactérie est un moyen de production largement utilisé par la communauté scientifique [2]. La surexpression de protéines dans la bactérie ne s'effectue cependant pas sans problèmes. En
20 effet, elle donne lieu à trois cas de figures :

- Le premier cas, idéal, est celui où la protéine est surproduite sous une forme correctement repliée dans l'espace lors de sa synthèse in vivo. Ce n'est pas une situation rare, mais elle n'est pas
25 fréquente non plus. Elle concerne essentiellement les protéines solubles et de taille réduite, c'est à dire d'environ 20 à 50 kDa.

- Le deuxième cas, le plus courant, est celui où la protéine est surproduite et agrégée sous la forme
30 de corps d'inclusion. Cela concerne les protéines polytopiques et/ou de grande taille. Dans ce cas, là

cinétique de repliement de la protéine est nettement plus lente que sa vitesse de biosynthèse. Cela favorise l'exposition au solvant aqueux des régions hydrophobes de la protéine qui sont normalement enfouies au coeur de celle-ci, et génère des interactions non spécifiques qui conduisent à la formation d'agrégats insolubles. Suivant le degré de désordre de ce repliement, les corps d'inclusions peuvent être solubilisés/dépliés dans des conditions non natives, avec de l'urée ou de la guanidine. La protéine solubilisée est ensuite soumise à différents traitements comme la dialyse ou la dilution pour favoriser, dans certains cas avec succès, un repliement 3D natif.

- Le troisième cas est celui où l'expression engendre une toxicité d'un degré variable. Celui-ci va de l'absence de produit d'expression si la bactérie parvient à s'adapter, à la mort de celle-ci si le produit est trop toxique. C'est un cas qui se présente assez fréquemment et le plus souvent avec des protéines ou des domaines de protéines membranaires, comme par exemple ceux des protéines d'enveloppe du virus de l'hépatite C [5] ou du virus d'immunodéficience humain [6].

Le problème de toxicité concerne essentiellement l'expression de protéines membranaires, c'est à dire de protéines présentant un domaine hydrophobe. Or ces protéines présentent un intérêt grandissant. Elles sont d'une part relativement nombreuses puisque l'établissement des différents génomes confirme qu'elles représentent environ 30% des protéines potentiellement codées par ces génomes [7].

Elles constituent d'autre part 70% des cibles thérapeutiques et leur altération est à l'origine de nombreuses maladies génétiques [8].

Il est donc essentiel de mettre au point des méthodes facilitant ou permettant l'expression de telles protéines ou de leur partie membranaire.

Des efforts ont été faits dans ce sens avec par exemple la mise au point de souches bactériennes, soit tolérant mieux l'expression de protéines membranaires [9, 10], soit ayant une régulation plus stricte du mécanisme d'expression comme dans le cas de la souche *E. coli* BL21(DE3)pLysS développée par Stratagene. Cependant, ces améliorations ne permettent pas de supprimer le phénomène de toxicité dans tous les cas, en particulier lors de l'expression de peptides hydrophobes correspondant à des ancres membranaires.

L'un des enjeux médicaux majeurs actuels concerne le traitement de l'hépatite C. Celle-ci est provoquée par le virus de l'hépatite C (VHC) de la famille des flaviviridae et qui infecte spécifiquement les cellules hépatiques [11]. Ce virus est constitué d'un ARN positif d'environ 9500 bases qui code pour une polyprotéine de 3033 résidus [13], symbolisé sur la figure 1 annexé par le rectangle 1A. Cette polyprotéine est clivée après expression par des protéases endogènes et exogènes pour donner naissance à 10 protéines différentes. Deux d'entre elles, appelées E1 et E2, sont glycosylées et forment l'enveloppe du virus. Elles possèdent chacune des domaines membranaires appelés TM, en particulier TME1 pour la protéine E1 et TME2 pour la protéine E2. Les positions de clivage qui les génèrent

sont indiquées sur la figure 1 par des flèches avec, mentionné au-dessous, un nombre qui correspond à la position dans la polyprotéine du premier acide aminé de séquence résultant du clivage. Les protéines E1 et E2 sont symbolisées par un rectangle. La partie blanche de chaque rectangle correspond à l'ectodomaine (ed) et la partie grisée à la région transmembranaire (TM). La séquence primaire des TM est indiquée dans le bas de la figure en code à une lettre, avec des nombres correspondant à la position des acides aminés dans la polyprotéine situés aux extrémités de ces domaines. Les étoiles indiquent les acides aminés hydrophobes. Ces domaines membranaires ou régions membranaires du virus ont des propriétés d'association particulières qui conditionnent la structuration de l'enveloppe du virus [12]. A ce titre, ils constituent des cibles thérapeutiques potentielles. La compréhension du mécanisme d'association du virus nécessite des études de la structure 3D de ces domaines notamment par les techniques précitées, ce qui implique de produire ces peptides en quantité abondante, et en plus de préférence par voie biosynthétique pour permettre un marquage isotopique ^{15}N et/ou ^{13}C .

Les différents essais d'expression de E1 de l'art antérieur, notamment dans *E. coli* [14][5] ou dans des cellules d'insecte *sf9* infectées par baculovirus [15], n'ont pas permis de surproduire cette protéine E1, notamment du fait de la toxicité induite par son expression, y compris dans les souches *E. coli* BL21(DE3)pLysS dites « résistantes » décrites ci-dessus. Il n'y a pas eu d'essai de surexpression de la

protéine E2 dans la bactérie. Ces problèmes de toxicité sont essentiellement dus à la région C-terminale des deux protéines, riche en acides aminés hydrophobes qui forment des domaines transmembranaires assurant l'ancrage à la membrane du réticulum endoplasmique.

Il existe donc un réel besoin pour un système d'expression de protéines toxiques qui ne présente pas les inconvénients, limitations, défauts et désavantages des techniques de l'art antérieur.

En outre, il existe un réel besoin d'un vecteur d'expression comprenant un tel système d'expression de protéines toxiques permettant de mettre en œuvre un procédé de fabrication de protéines toxiques qui ne présente pas les inconvénients, limitations, défauts et désavantages des techniques de l'art antérieur.

EXPOSÉ DE L'INVENTION

Le but de la présente invention est précisément de fournir un système d'expression d'une protéine toxique, qui réponde, entre autres, aux besoins indiqués ci-dessus.

Ce but, et d'autres encore, sont atteints, conformément à l'invention, par un système d'expression caractérisé en ce qu'il comprend successivement, dans le sens 5'-3', une séquence nucléotidique codant pour le dipeptide Asp-Pro, appelée ci-dessous séquence *dp*, et une séquence nucléotidique (*pt*) codant pour une protéine toxique (Pt). Ce système sera identifié ci-dessous par : *dp-pt*.

Selon un mode particulièrement préféré de la présente invention, le système d'expression comprend en

outre, en amont de la séquence *dp*, une séquence nucléotidique (*ps*) codant pour une protéine soluble (*Ps*). Cette protéine soluble peut être par exemple la glutathion S-transférase (GST) ou la thiorédoxine (TrX) ou une autre protéine soluble équivalente. Ce système d'expression selon l'invention sera identifié ci-dessous par : *ps-dp-pt*.

Le système d'expression *dp-pt* de la présente invention, qui comprend une séquence codant pour Asp-Pro (DP en code à une lettre) placée en amont de la séquence nucléotidique de la protéine toxique, permet de manière tout à fait inattendue de supprimer l'effet toxique de la protéine pour la cellule hôte. En outre, les inventeurs ont noté que, de manière tout à fait surprenante, la suppression de toxicité de la protéine chez l'hôte est encore plus efficace avec le système d'expression *ps-dp-pt*, lorsque le peptide toxique est produit en fusion C-terminale avec une protéine soluble, par exemple la glutathion S-transférase ou la thiorédoxine, avec la séquence Asp-Pro insérée entre la protéine soluble et le peptide toxique.

Le système d'expression *dp-pt* ou *ps-dp-pt* de la présente invention permet la surproduction de protéines toxiques dans des cellules hôtes, en particulier de protéines hydrophobes, notamment de peptides qui correspondent à, ou qui comprennent, des domaines hydrophobes de protéines ancrées aux membranes. Il peut s'agir par exemple d'une protéine membranaire ou un domaine d'une protéine membranaire. Il peut s'agir par exemple d'une protéine d'un virus, par exemple d'un virus de l'hépatite C, d'un virus du sida, ou de tout

autre virus pathogène pour l'humain et de manière générale pour les mammifères.

Par exemple, le système *dp-pt* ou *ps-dp-pt* de l'invention permet de surproduire chez un hôte tel que *E.coli* les domaines transmembranaires des protéines E1 et E2 du virus de l'hépatite C, appelées TME1 et TME2, correspondant respectivement aux séquences :

TME1 : 347-MIAGAHWGVLAGIAYFSMVGNWAKVLVLLLFAGVDA-383
10 Séquence IDn°1

TME2 : 717-MEYVLLFLLLADARVCCLWMMLLISQAEA-746
Séquence IDn°2

15 alors que cela n'était pas possible avec les techniques de l'art antérieur.

Les séquences nucléotidiques utilisables pour constituer le système *dp-pt* de l'invention codant pour les protéines TME1 (*dp-pt*_(TME1)) ou TME2 (*dp-pt*_(TME2))
20 peuvent être toutes les séquences possibles codant respectivement pour les protéines de fusion DP-TME1 et DP-TME2. Les séquences codant pour les protéines TME1 et TME2 peuvent être par exemple, avantageusement, respectivement, des séquences IDn°3 et IDn°4 de la
25 liste de séquences annexée. Pour obtenir le système *dp-pt* la séquence *dp* codant pour le dipeptide Asp-Pro (DP) est ajoutée à ces séquences.

Les séquences nucléotidiques utilisables pour constituer le système *ps-dp-pt* de l'invention codant
30 pour les protéines TME1 (*ps-dp-pt*_(TME1)) ou TME2 (*ps-dp-pt*_(TME2)) peuvent être toutes les séquences possibles

codant respectivement pour les protéines de fusion Ps-DP-TME1 et Ps-DP-TME2. Il peut s'agir par exemple, avantageusement, des séquences IDn°34, IDn°35 et IDn°36 de la liste de séquences annexée pour TME1, permettant
5 d'obtenir une protéine chimère Ps-DP-TME1. Il peut s'agir par exemple, avantageusement, des séquences IDn°37, IDn°38 et IDn°39 de la liste de séquences annexée pour TME2, permettant d'obtenir une protéine chimère Ps-DP-TME2.

10 En effet, les séquences nucléotidiques précitées présentent des codons optimisés pour l'expression de TME1 et TME2 dans une bactérie, par exemple chez *E. coli*.

Il existe un grand nombre de séquences de l'ARN du
15 VHC produisant un phénotype infectieux : ces séquences sont également utilisables dans la présente invention.

La séquence codant pour le dipeptide Asp-Pro peut être par exemple : gacccg, ou toute autre séquence codant pour ce dipeptide.

20 La séquence codant pour la GST peut être par exemple celle présente dans les plasmides pGEXKT dont la séquence correspond à la séquence IDn°29 de la liste de séquence annexée ou toute séquence équivalente, c'est à dire codant pour cette protéine soluble. La séquence codant pour
25 la TrX peut être par exemple celle présente dans le plasmide d'expression pET32a+ dont la séquence correspond à la séquence IDn°30 de la liste de séquence annexée ou toute séquence équivalente, c'est à dire codant pour cette protéine soluble.

30 Pour la fabrication de la protéine toxique, le système d'expression dp-pt ou ps-dp-pt de l'invention est placé à l'intérieur d'une cellule hôte par exemple par clonage dans un plasmide approprié et ceci par les

techniques habituelles de transformation d'un hôte dans les techniques de recombinaison génétique.

Le plasmide dans lequel le système d'expression de la présente invention peut être cloné pour former ce vecteur sera choisi notamment en fonction de la cellule hôte. Il peut s'agir par exemple du plasmide pT7-7 (séquence IDn°33 de la liste de séquences annexée), d'un plasmide de la série pGEX (par exemple de séquence IDn°31 de la liste de séquences annexée), commercialisé par exemple par la société Pharmacia, ou un plasmide de la série pET32 (par exemple de séquence IDn°32 de la liste de séquences annexée), commercialisé par exemple par la société Novagen.

Les plasmides de la série pGEX et de la série pET32 seront avantageusement utilisés pour la mise en œuvre de la présente invention. En effet, ils comprennent déjà une séquence *ps* codant pour une protéine soluble (*Ps*), respectivement pour la glutathion S-transférase et pour la thiorédoxine. Aussi, avantageusement, le système *dp-pt* sera cloné dans ces plasmides en aval de cette séquence *ps* codant pour la protéine soluble.

La présente invention se rapporte donc également à un vecteur d'expression comprenant un système d'expression selon l'invention *dp-pt* ou *ps-dp-pt* ; en particulier un vecteur comprenant un système d'expression *dp-pt* selon l'invention et la séquence oligonucléotidique du plasmide pT7-7, ou un vecteur comprenant un système d'expression *ps-dp-pt* selon l'invention et la séquence oligonucléotidique d'un plasmide pGEX ou d'un plasmide pET32.

Par exemple, les vecteurs d'expression de la présente invention, convenant à un hôte bactérien tel que *E. coli*, et permettant la surexpression de la protéine membranaire TME1 précitée, peuvent avoir
5 avantageusement une séquence oligonucléotidique choisie parmi les séquences IDn°40 (avec pGEXKT), IDn°42 (avec pET32a+) et IDn°44 (avec pT7-7) de la liste de séquences annexée.

Par exemple, les vecteurs d'expression de la
10 présente invention, convenant à un hôte bactérien tel que *E. coli*, et permettant la surexpression de la protéine membranaire TME2 précitée, peuvent avoir avantageusement une séquence oligonucléotidique choisie
15 parmi les séquences IDn°41 (avec pGEXKT), IDn°43 (avec pET32a+) et IDn°45 (avec pT7-7) de la liste de séquences annexée.

En effet, les vecteurs d'expression précités présentent des codons optimisés pour l'expression des protéines chimères de la présente invention, incluant
20 TME1 et TME2, dans une bactérie, par exemple chez *E. coli*.

La présente invention se rapporte également à une cellule procaryote transformée par un vecteur d'expression selon l'invention. Cette cellule
25 procaryote transformée par le vecteur d'expression de la présente invention doit de préférence permettre la surexpression de la protéine toxique pour laquelle code le vecteur. Ainsi, toute cellule hôte capable de d'exprimer le vecteur d'expression de la présente
30 invention est utilisable, par exemple *E. coli*, avantageusement, la souche *E. coli* BL21(DE3)pLysS.

La présente invention se rapporte également à un procédé de fabrication d'une protéine toxique par recombinaison génétique comprenant les étapes suivantes :

- 5 - transformer une cellule hôte avec un vecteur d'expression selon l'invention,
- cultiver la cellule hôte transformée dans des conditions de culture telles qu'elle fabrique une protéine de fusion comprenant le dipeptide Asp-Pro
- 10 suivi de la séquence peptidique de la protéine toxique à partir dudit vecteur d'expression, et
- isoler ladite protéine de fusion, et
- cliver ladite protéine de fusion pour récupérer la protéine toxique.

15 Les étapes de transformation, de culture, ainsi que de l'isolement de la protéine chimère fabriquée peuvent être réalisées par les techniques habituelles de recombinaison génétique, par exemple par des techniques telles que celles qui sont décrites dans le

20 document [25].

L'étape consistant à isoler la protéine de fusion peut être réalisée par les techniques habituelles connues de l'homme du métier pour isoler une protéine d'un extrait cellulaire.

25 La protéine de fusion fabriquée par le procédé de l'invention a une séquence « Protéine soluble-Asp-Pro-protéine toxique ». Dans la présente description, le dipeptide Asp-Pro est aussi appelé DP suivant le code à une lettre des acides aminés.

30 Par exemple, lorsque la protéine toxique est TME1, la protéine de fusion peut avoir la séquence

IDn°46 de la liste de séquences annexée, qui correspond à la protéine de fusion GST-DP-TME1 ; la séquence IDn°48 de la liste de séquences annexée, qui correspond à la protéine de fusion TrX-DP-TME1 ; ou la séquence
5 IDn°50 de la liste de séquences annexée, qui correspond à la protéine de fusion M-DP-TME1 de la liste de séquences annexée.

Par exemple, lorsque la protéine toxique est TME2, la protéine de fusion peut avoir la séquence
10 IDn°47 de la liste de séquences annexée, qui correspond à la protéine de fusion GST-DP-TME2 ; la séquence IDn°49 de la liste de séquences annexée, qui correspond à la protéine de fusion TrX-DP-TME2 ; ou la séquence IDn°51 de la liste de séquences annexée, qui correspond
15 à la protéine de fusion M-DP-TME2 de la liste de séquences annexée.

L'étape de clivage de cette protéine de fusion peut être réalisée avantageusement au moyen d'acide formique qui clive la protéine de fusion au niveau du
20 dipeptide Asp-Pro. Elle peut être réalisée par ailleurs par toute technique appropriée connue de l'homme du métier pour récupérer une protéine d'un échantillon à partir d'une protéine de fusion.

25 Les inventeurs sont les premiers à avoir trouvé un système réellement efficace pour produire et même de surproduire, notamment dans la bactérie *Escherichia coli* (*E. coli*) des peptides hydrophobes correspondant aux domaines membranaires des protéines E1 et E2 de
30 l'enveloppe du virus de l'hépatite C, dont l'expression est létale pour le microorganisme.

Le champ d'applications de la présente invention concerne principalement la production de peptides hydrophobes à grande échelle notamment pour la recherche fondamentale et industrielle. En outre, la production de la protéine chimère constituée de la protéine soluble, du dipeptide Asp-Pro, et du peptide hydrophobe peut être utilisée dans un but fonctionnel, notamment pour renseigner sur le degré d'oligomérisation du domaine membranaire ou encore sur sa capacité d'hétéropolymérisation.

Les protéines de fusion, ou protéines chimères, sont produites via leur ADN codant présent par exemple dans des plasmides commerciaux et à la suite desquels est introduit en phase l'ADN codant pour la séquence Asp-Pro suivi de celui codant pour le peptide toxique. Cette application peut être commercialisée sous la forme de plasmides d'expression bactériens qui incluront la séquence du site Asp-Pro, en aval de celle des protéines solubles déjà présentes. Le plasmide correspondant sera décrit par exemple comme un outil facilitant la production par voie biologique de peptides ou protéines membranaires toxiques.

Ainsi, la présente invention est applicable à tout système de surexpression de protéines recombinantes, avec ou sans fusion à une protéine soluble comme par exemple la GST, ou la thiorédoxine, incluant une séquence Asp-Pro non naturelle insérée en amont d'une séquence codant pour un domaine toxique de la protéine, par exemple un domaine membranaire d'une protéine.

D'autres caractéristiques et avantages de la présente invention apparaîtront encore à l'homme du métier à la lecture des exemples suivants donnés à titre illustratif et non limitatif en référence à la
5 liste de séquences et aux figures annexées.

Brève description de la liste de séquences annexée

- Séquences IDn°1 et 2 : respectivement, séquences peptidiques de TME1 et de TME2.
- 10 - Séquences IDn°3 et 4 : respectivement, séquences codant pour le peptide TME1 et le peptide TME2.
- Séquences IDn°5 et 6 : respectivement, oligonucléotide (+) d'insertion dans pT7-7 (OL13(+)) et oligonucléotide (-) d'insertion dans
15 pT7-7 (OL14(-)).
- Séquences IDn°7 et 8 : respectivement, ADN sens codant de TME1 + site *cl*a I en 3' et ADN sens anticodant de TME1 + site *cl*a I en 5' (séquence complémentaire de la séquence IDn°7).
- 20 - Séquences IDn°9 et 10 : respectivement, oligonucléotide sens codant (OL11(+)) et oligonucléotide sens anticodant (OL12(-)) pour la synthèse de TME1.
- Séquence IDn°11 : oligonucléotide (+)
25 d'insertion dans pGEXKT sans site *dp* (OL15(+)).
- Séquence IDn°12 : oligonucléotide (+)
d'insertion dans pGEXKT avec site *dp* (OL17(+)).
- Séquence IDn°13 : oligonucléotide (-)
d'insertion dans pGEXKT (OL16(-)).
- 30 - Séquence IDn°14 : oligonucléotide (+)
d'insertion dans pET32a (OL18(+)) (s'hybride au segment 915-932 de pGEXKT).

- Séquences IDn°15 et 16 : respectivement, oligonucléotides (+) (OL19(+)) et (-) (OL20(-)) d'insertion dans pT7-7 de l'ADN codant pour MDP-TME1.
- 5 - Séquences IDn°17 et 18 : respectivement, oligonucléotide (+) d'insertion dans pT7-7 (OL23(+)) et oligonucléotide (-) d'insertion dans pT7-7 (OL24(-)) .
- Séquences IDn°19 et 20 : respectivement, ADN sens codant pour TME2 + site de *Nde* I en 5' et *Hind* III en 3' ; et ADN sens anticodant de TME2 + site *Nde* I en 3' et *Hind* III en 5' (séquence complémentaire de IDn°17).
- 10 - Séquences IDn°21 et 22 : respectivement, oligonucléotides sens codant (OL21(+)) et anticodant (OL22(-)) pour la synthèse de TME2.
- 15 - Séquence IDn°23 : oligonucléotide (+) d'insertion dans pGEXKT sans site dp (OL25(+)).
- Séquences IDn°24 et 25 : respectivement, oligonucléotides (+) (OL27(+)) et (-) (OL26(-)) d'insertion dans pGEXKT avec site dp.
- 20 - Séquences IDn°26 et 27 : respectivement, oligonucléotides (+) (OL28(+)) et (-) (OL29(-)) d'insertion dans pT7-7 de l'ADN codant pour MDP-TME2.
- 25 - Séquence IDn°28 : fin de la séquence de la protéine soluble GST suivie du site thrombine codée dans le plasmide pGEXKT.
- Séquence IDn°29 : ADN codant pour la protéine GST dans le plasmide pGEXKT.
- 30

- Séquence IDn°30 : ADN codant pour la thiorédoxine (TrX) dans le plasmide pET32a+.
- Séquences IDn°31, 32 et 33 : respectivement, plasmides d'expression pGEXKT, pET32a+ et pT7-7.
- 5 - Séquences IDn°34, 35 et 36 : respectivement, systèmes d'expression selon l'invention codant pour les protéines de fusion GST-DP-TME1, TrX-DP-TME1 et M-DP-TME1.
- 10 - Séquences IDn°37, 38 et 39 : respectivement, systèmes d'expression selon l'invention codant pour les protéines de fusion GST-DP-TME2, TrX-DP-TME2 et M-DP-TME2.
- Séquences IDn°40 et 41 : respectivement, vecteurs d'expression pGEXKT-dp-pt_{TME1} et pGEXKT-dp-pt_{TME2} selon l'invention codant pour les protéines de fusion GST-DP-TME1 et GST-DP-TME2.
- 15 - Séquences IDn°42 et 43 : respectivement, vecteurs d'expression pET32a-dp-pt_{TME1} et pET32a-dp-pt_{TME2} selon l'invention codant pour les protéines de fusion TrX-DP-TME1 et TrX-DP-TME2 (codent par le brin complémentaire).
- 20 - Séquences IDn°44 et 45 : respectivement, vecteurs d'expression pT7-7-dp-pt_{TME1} et pT7-7-dp-pt_{TME2} selon l'invention codant pour les protéines de fusion M-DP-TME1 et M-DP-TME2.
- 25 - Séquences IDn°46 et 47 : respectivement, protéines de fusion GST-DP-TME1 et GST-DP-TME2 selon l'invention obtenues à partir des plasmides pGEXKT-dp-pt_{TME1} et pGEXKT-dp-pt_{TME2}.
- 30 - Séquences IDn°48 et 49 : respectivement, protéines de fusion TrX-DP-TME1 et TrX-DP-TME2 selon

l'invention obtenues à partir des plasmides pET32a-dp-pt_{TME1} et pET32a-dp-pt_{TME2}.

- Séquences IDn°50 et 51 : respectivement, protéines de fusion M-DP-TME1 et M-DP-TME2 selon l'invention
5 obtenues à partir des plasmides pT7-7-dp-pt_{TME1} et pT7-7-dp-pt_{TME2}.
- Séquences IDn°52 et 53 : respectivement, protéines GST et TrX codées par le vecteur pGEXKT et pET32a+.

10 Brève description des figures

- Figure 1 : représentation schématique d'une partie
de la polyproptéine du VHC et séquence peptidique des
domaines membranaires C-terminaux des protéines
d'enveloppe E1 et E2. Les séquences peptidiques
15 représentées correspondent au type infectieux #D00831
et #M67463 pour TME1 et TME2 respectivement obtenus de
la banque publique de séquences Du Laboratoire Européen de
Biologie Moléculaire (« European Molecular Biology
Laboratory ») (EMBL).
- 20 - Figure 2 : création de l'ADN codant pour le domaine
membranaire C-terminal de la protéine d'enveloppe E1 du
VHC et séquences additionnelles en 5' et 3' pour
clonage dans différents plasmides. Les séquences
représentées sur cette figure sont reportées dans la
25 liste de séquences annexée.
- Figure 3 : création de l'ADN codant pour le domaine
membranaire C-terminal de la protéine d'enveloppe E2 du
VHC et séquences additionnelles en 5' et 3' pour
clonage dans différents plasmides. Les séquences
30 représentées sur cette figure sont reportées dans la
liste de séquences annexée.

- Figure 4, panneaux A à F : toxicité des domaines membranaires exprimés dans la bactérie et suppression de cette toxicité par insertion d'un site dp. Les panneaux A, C et E sont des représentations graphiques de mesures de la densité optique (DO) à 600 nm en fonction du temps (t) en heures de la production de différentes protéines dans une bactérie utilisant ou non le système d'expression de la présente invention. Les panneaux B, D et F sont des représentations des gels de migration des protéines respectivement des panneaux A, C et E.

- Figures 5 A et B : surexpression des protéines chimères thiorédoxine-Asp-Pro-Pt (Pt = domaines membranaires des protéines) dans la bactérie. La figure 5A est une représentation graphique de mesures de la densité optique (DO) à 600 nm en fonction du temps en heures de la production de différentes protéines dans une bactérie utilisant ou non le système d'expression de la présente invention. La figure 5B est une représentation d'un gel de migration des protéines de la figure 5A.

- Figure 6 : expression et purification de la protéine de fusion (ou chimère) GST-TME2, et comparaison avec la GST seule. Cette figure représente en haut les séquences peptidiques de GST et GST-TME2, et en bas les gels obtenus par électrophorèse montrant que contrairement à GST seule, GST-TME2 est insoluble. Cette dernière est produite sous forme de corps d'inclusion qui ne peuvent pas se replier correctement.

- Figures 7A et 7B : représentations graphiques de résultats expérimentaux comparatifs montrant l'effet du

dipeptide DP (séquence oligonucléotidique *dp-pt* conforme à la présente invention) et du dipeptide DP et de la protéine soluble (séquence oligonucléotidique *ps-dp-pt* conforme à la présente invention) sur la synthèse
5 des protéines toxiques TME1 et de TME2 conformément à la présente invention.

EXEMPLES

10 Dans ces exemple, les oligonucléotides utilisés ont été commandés aux Laboratoires EUROBIO (<http://www.eurobio.fr/>); les plasmides ont été préparés avec le kit QIAprep (marque de commerce) de QIAGEN (<http://www.qiagen.com/>); les séquences d'ADN
15 ont été séquencées avec le kit ABI PRISM (marque déposée) BigDye (marque de commerce) Terminator cycle d'Applied Biosystems (<http://home.appliedbiosystems.com/>); les souches d'*E. coli* BL21(DE3) et BL21(DE3)pLysS ont été obtenues chez
20 Stratagene (<http://www.stratagene.com/>); les souches C41 et C43 (BL21(DE3)) ont été fournies par le Dr. Bruno Miroux (CNRS-CEREMOD, Centre de recherches sur l'endocrinologie moléculaire et le développement ; les enzymes de restriction et modification d'ADN ont été
25 obtenues chez New England Biolabs (<http://www.neb.com/neb/>); les électrophorèses de protéines ont été réalisées avec un miniprotean 3 (marque de commerce) des Laboratoires Bio-Rad (<http://www.bio-rad.com/>); le plasmide pCR (marque
30 déposée) T7 topo TA a été obtenu chez Invitrogen (<http://www.invitrogen.com/>); le plasmide pET32a+ a été

obtenu chez Novagen (<http://www.novagen.com>); les plasmides pT7-7 et pGP1-2 et la souche K38 [22] ont été demandés au Prof. Tabor (Department of Biological Chemistry, Harvard Medical School); le plasmide pGEX-KT a été demandé au Prof. Dixon (Department of Biological Chemistry, University of Michigan Medical School); les autres produits ont été obtenus chez Sigma (<http://sigma.aldrich.com>).

Dans les exemples suivants, la production des peptides TME1 et TME2 a été dans un premier temps réalisée sans le système d'expression de la présente invention, puis en fusion avec une protéine soluble, et enfin en fusion avec la GST avec insertion du site Asp-Pro (« DP » en codage à une lettre) entre la protéine soluble et TME1 ou TME2.

L'abréviation « SEQ IDn° » est utilisée pour « séquence IDn° » et renvoie à la liste de séquences annexée.

Exemple 1 : Synthèse du système d'expression

1.1) CONSTRUCTION DES VECTEURS D'EXPRESSION pT7-7-pt_{TME1} ET pT7-7-pt_{TME2}

L'ADN codant pour les deux domaines a été synthétisé de novo à partir des oligonucléotides appropriés. Les codons ont été choisis suivant leur plus grande fréquence d'utilisation dans la bactérie, telle qu'elle a été quantifiée par Sharp et col. [17]. Les constructions sont décrites sur la figure 2 annexée pour TME1 et sur la figure 3 annexée pour TME2.

Chaque ADN synthétique a été généré en utilisant un jeu de deux oligonucléotides longs et

chevauchants, OL11 (SEQ IDn°9) et OL12 (SEQ n°10) pour TME1 et OL21 (SEQ IDn°19) et OL22 (SEQ IDn°20) pour TME2, qui ont été amplifiés après hybridation avec deux oligonucléotides externes choisis en fonction du
5 clonage dans un plasmide donné. Ainsi les clonages dans pT7-7 ont été réalisés en utilisant le jeu d'oligonucléotides externes OL13 (SEQ IDn°5) et OL14 (SEQ IDn°6) pour TME1 et OL23 (SEQ IDn°15) et OL24 (SEQ IDn°16) pour TME2.

10 Chaque ADN synthétique a été généré en utilisant un jeu de quatre oligonucléotides : deux longs et chevauchants et deux courts et externes. Les ADN ont été amplifiés par la méthode de réaction de polymérisation en chaîne (« polymérase chain
15 reaction ») dite de "PCR" [18] puis clonés dans un plasmide bactérien pCR (marque de commerce) T7 topo TA. Les ADN synthétisés ont été séquencés puis sous-clonés dans le vecteur d'expression bactérien pT7-7 [19] en utilisant les sites de restriction *Nde* I en 5' et *Cla* I
20 ou *Hind* III en 3'.

Sur la figure 2 :

A : Séquence peptidique TME1 du sous-type #D00831. La numérotation correspond à la position de la séquence
25 dans la polyprotéine comme décrit sur la figure 1.

B : Séquence d'ADN codant pour le domaine membranaire avec des codons optimisés pour l'expression dans la bactérie.

C et D : Stratégie d'amplification sans matrice de
30 l'ADN. Le sens codant et le sens anticodant des oligonucléotides sont indiqués respectivement par les

signes (+) et (-). Les oligonucléotides longs sont chevauchants d'une vingtaine de bases pour créer l'amorce puis la matrice. Les oligonucléotides courts permettent d'amplifier la matrice par PCR en intégrant les sites de restriction désirés et suivant les plasmides utilisés. L'insertion dans pT7-7 a été réalisée avec le couple d'oligonucléotides OL13 (SEQ IDn°5) et OL14 (SEQ IDn°6), via un sous-clonage dans pCRT7 topo, en intégrant les sites *Nde* I et *Hind* III.

10 L'insertion dans pGEXKT a été effectuée suivant la même méthode avec le couple d'oligonucléotides OL15 (SEQ IDn°11) et OL16 (SEQ IDn°13) en intégrant les sites *Bam*H I et *Eco*R I. L'insertion du site dp (gacccg) et le clonage dans pGEXKT ont été effectués avec le couple

15 d'oligonucléotides OL17 (SEQ IDn°12) et OL16 (SEQ IDn°13). La construction dans pGEXKT a été transférée dans pET32a qui code pour la thiorédoxine avec le couple d'oligonucléotides OL18 (SEQ IDn°14) et OL16 (SEQ IDn°13). L'oligonucléotide OL18 (SEQ IDn°14)

20 s'hybride dans la région terminale de l'ADN codant pour la GST dans pGEXKT. La séquence amplifiée intègre la fin de la GST (SDLSGGGGG) suivie du site thrombine (LVPRGS) (SEQ IDn°28), du site DP et du passage membranaire. Après clonage, l'ADN inséré dans pET32a

25 permet d'exprimer la chimère thiorédoxine-SDLSGGGGGLVPRGS-DP-TME1 (SEQ IDn°48).

Sur la figure 3 :

La légende est identique à la figure 2, mais la

30 séquence peptidique est celle du sous-type #M67463. L'insertion dans pT7-7 a été réalisée avec le couple

d'oligonucléotides OL23 et OL24 (respectivement SEQ IDn°17 et SEQ IDn°18) via un sous-clonage dans pCRT7 topo en intégrant les sites *Nde* I et *Hind* III.

L'insertion dans pGEXKT a été effectuée suivant
5 la même méthode avec le couple d'oligonucléotides OL25 et OL26 (respectivement SEQ IDn°23 et SEQ IDn°25) intégrant les sites *Bam*H I et *Eco*R I. L'insertion du site *dp* (gacccg) et le clonage dans pGEXKT ont été effectués avec le couple d'oligonucléotides OL27 et
10 OL26 (respectivement SEQ IDn°24 et SEQ IDn°25). L'insertion dans pET32a a été effectuée comme décrit sur la figure 2 en utilisant le couple d'oligonucléotides OL18 et OL26 (respectivement SEQ IDn°14 et SEQ IDn°25).

15

1.2) CONSTRUCTION DES VECTEURS D'EXPRESSION
pGEXKT-pt_{TME1}, pGEXKT-pt_{TME2}, pGEXKT-dp-pt_{TME1} ET
pGEXKT-dp-pt_{TME2}

Les vecteurs d'expression pGEXKT-pt_{TME1} et
20 pGEXKT-pt_{TME2} ont été construits par PCR comme décrits sur les figures 2 et 3 annexées. L'ADN-matrice utilisé pour amplifier les ADN codant pour TME1 ou TME2 est celui cloné dans les plasmides pT7-7. Le clonage dans le plasmide pGEXKT [20, 21] de TME1 a été réalisé en
25 utilisant les jeux d'oligonucléotides OL15 (SEQ IDn°11) et OL16 (SEQ IDn°13) permettant l'insertion des sites de restriction *Bam*H I en 5' et *Eco*R I en 3'. Le clonage de TME2 dans le même vecteur a été réalisé en utilisant les jeux d'oligonucléotides OL25 (SEQ IDn°21) et OL26
30 (SEQ IDn°23).

Comme indiqué dans la figure 2, l'insertion du site *dp* en position N-terminale de TME1 a été réalisée en remplaçant l'oligonucléotide 5' OL15 (SEQ IDn°11) par l'oligonucléotide OL17 (SEQ IDn°12). L'insertion du site *dp* en position N-terminale de TME2 a été réalisée en remplaçant l'oligonucléotide 5' OL25 (SEQ IDn°21) par l'oligonucléotide OL27 (SEQ IDn°22), comme montré dans la figure 3.

10 1.3) CONSTRUCTION DES VECTEURS D'EXPRESSION
 pET32a-dp-TME1 ET pET32a-dp-TME2

Les vecteurs d'expression pET32a-dp-TME1 et pET32a-dp-TME2 ont été construits par PCR comme décrits sur les figures 2 et 3 annexées, en utilisant le jeu d'oligonucléotides indiqué. L'oligonucléotide amont intègre un site *EcoR* V et s'hybride avec la région terminale de gène codant pour la GST. Il permet d'intégrer la queue de 5 glycines et le site de clivage par la thrombine présents dans le plasmide. L'oligonucléotide en aval est le même que celui utilisé pour le clonage dans pGEXKT.

L'insertion dans le plasmide pET32a se fait par les sites *Msc* I/*EcoR* V en 5' et *EcoR* I en 3'. Il permet d'insérer en phase à la fin de la séquence de la thiorédoxine la queue de 5 glycines, le site de clivage par la thrombine, le site DP et le passage membranaire. Le plasmide d'origine pET32a, qui sert de témoin, code pour la thiorédoxine suivie d'une séquence intégrant divers éléments qui n'ont pas été supprimés et qui contribuent pour une bonne part à la masse de la protéine chimère fabriquée.

L'ADN-matrice utilisé pour amplifier les ADN codant pour TME1 ou TME2 est celui cloné dans les plasmides pGEXKT-*dp-pt*_{TME1} ou pGEXKT-*dp-pt*_{TME2}. Pour TME1, le clonage dans pET32a+ a été réalisé en utilisant les jeux d'oligonucléotides OL18 (SEQ IDn°14) et OL16 (SEQ IDn°13). Le clonage de TME2 dans le même vecteur a été réalisé en utilisant les jeux d'oligonucléotides OL18 (SEQ IDn°14) et OL26 (SEQ IDn°23), comme indiqué dans la figure 3.

10.

Exemple 2 : Expression de séquences codant pour les protéines TME1 et TME2 seules

L'expression des séquences codant pour les domaines seuls TME1 et TME2 a été testée par induction thermique ou chimique et en utilisant différentes souches bactériennes comme décrit ci-après.

15

2.1) SYSTEME D'INDUCTION THERMIQUE

Le système développé par Tabor [22] permet d'exprimer une protéine par induction thermique en utilisant deux vecteurs dans une même bactérie, pT7-7 et pGP1-2.

20

Le plasmide pT7-7 contient l'ADN à exprimer, placé sous contrôle d'un promoteur $\phi 10$ reconnu par l'ARN polymérase du phage T7. Le plasmide pGP1-2 contient le gène codant pour la polymérase du phage T7, placé sous le contrôle d'un promoteur λp_L . Ce promoteur est réprimé par un répresseur thermosensible, cI857, présent lui aussi dans pGP1-2. A 30°C, cI857 est normalement exprimé et réprime le promoteur λp_L , ce qui

25

30

bloque l'expression de la polymérase et donc aussi celle de la protéine d'intérêt.

L'induction est déclenchée par un passage de la culture de 37 à 42°C pendant 15-30 min, puis
5 l'expression se poursuit à 37°C. Ce système est donc particulièrement bien adapté lorsqu'il est nécessaire de contrôler strictement l'expression d'une protéine donnée, notamment si celle-ci est toxique pour la bactérie.

10

2.2) SYSTEME D'INDUCTION CHIMIQUE

Le même plasmide pT7-7 contenant l'ADN à exprimer est cette fois introduit dans des bactéries *E. coli* de type BL21(DE3) (B F⁻ dcm ompT hsdS(r_B⁻m_B⁻) gal λ
15 (DE3) et BL21(DE3)pLysS (B F⁻ dcm ompT hsdS(r_B⁻m_B⁻) gal λ (DE3) [pLysS Cam^r]). Ces bactéries ont été modifiées pour contenir dans le génome une copie du gène codant pour l'ARN polymérase du phage T7, placé sous contrôle d'un promoteur lacUV5 inductible par l'isopropyle-1-
20 thio-β-D-galactoside (IPTG). Dans ce cas, les bactéries sont cultivées à leur température optimum de 37°C ou moins si nécessaire. L'expression est induite par addition d'IPTG dans la culture. La souche BL21(DE3)pLysS est particulièrement bien adaptée aux
25 protéines dont l'expression basale est toxique pour la bactérie hôte. En effet, la présence du plasmide pLysS permet l'expression continue et à un faible niveau de lysozyme du phage T7. Celui-ci inhibe la polymérase du phage T7, dont la faible expression en absence
30 d'induction pourrait permettre l'expression basale de protéine toxique.

Les inventeurs ont aussi testé l'expression des domaines membranaires seuls dans des souches appelées C41 et C43 [10] qui ont été sélectionnées pour résister à l'expression de protéines membranaires toxiques. Ces souches sont dérivées de la souche BL21(DE3) et sont utilisées de la même façon que cette dernière.

2.3) TESTS D'EXPRESSION

Suivant le système testé, les plasmides correspondants ont été introduits par transformation dans les différentes souches d'*E. coli* : K38 (HfrC λ) pour le système d'induction thermique de Tabor ou les différentes souches BL21 pour l'induction chimique. Le tableau 1 suivant résume les tests effectués.

Tableau 1

Induction	Souche	Plasmide
Thermique	K38	pT7-7 + pGP1-2
Chimique	BL21(DE3)	pT7-7
Chimique	BL21(DE3)pLysS	pT7-7
Chimique	C41 (BL21(DE3))	pT7-7
Chimique	C43 (BL21(DE3))	pT7-7

Dans chaque cas, une dizaine de transformants ont été mis en culture pour tester l'expression. Brièvement, les bactéries ont été mises en culture dans 5 ml de LB (10 g tryptone, 5 g extrait de levure, 5 g NaCl, qsp 1 litre H₂O) complémenté avec 50 μ g/ml d'ampicilline (nécessaire pour maintenir pT7-7 dans la bactérie) et 60 μ g/ml de kanamycine (nécessaire pour

maintenir pGP1-2 dans la bactérie) puis cultivées jusqu'à saturation soit à 30°C pour K38, soit à 37°C pour BL21(DE3). Les cultures sont ensuite diluées au 1/10 dans le même milieu de culture et cultivées
5 jusqu'à une densité optique (DO) de 1 mesurée à 600 nm sur un spectrophotomètre Philips PU8740 (marque de commerce).

L'expression est ensuite induite soit thermiquement (K38) à 42°C pendant 15 min, soit
10 chimiquement (BL21(DE3)) par addition de 1 mM d'IPTG. Elle est poursuivie pendant 3-5 heures à 37°C. La DO_{600nm} des cultures est mesurée à différents temps.

A la fin de l'expression, un volume de culture contenant l'équivalent de 0,1 DO de bactéries est
15 prélevé. Les bactéries sont récoltées par centrifugation et suspendues dans 50 µl de solution de lyse (SL : Tris-Cl 50 mM, pH 8,0, EDTA 2,5 mM, SDS 2%, urée 4 M, β-mercaptoéthanol 0,7 M). Après quelques minutes à température ambiante, 10 µl sont déposés sur
20 gel de polyacrylamide 16,5% pour électrophorèse de type "Tricine" [23] qui permet de bien séparer les protéines de petite masse moléculaire.

Sur la figure 4 :

25 Panneaux A, C et E : Les bactéries ont été transformées avec les plasmides pT7-7, pT7-7-TME1, pT7-7-TME2 (panneau A), pGEXKT, pGEXKT-TME1, pGEXKT-TME2 (panneau C) et pGEXKT-dp-TME1, pGEXKT-dp-TME2 (panneau E) puis
30 mises en culture et induites comme décrit ci-dessus. La croissance bactérienne a été suivie en mesurant l'augmentation de turbidité de chaque culture par

mesure de la densité optique à 600 nm en fonction du temps en heures.

Panneaux B, D, F : Les bactéries ont été prélevées au temps indiqué dans le texte et traitées comme décrit
5 ci-dessus. Elles ont ensuite été déposées sur gel d'électrophorèse, soit 16,5% d'acrylamide de type "Tricine" (panneau B), soit 14% d'acrylamide de type Laemmli SDS-PAGE (panneaux D et F). L'électrophorèse
montrée dans le panneau F a migré plus longtemps que
10 celle montrée dans le panneau D, ceci pour améliorer la séparation des bandes dans la zone des 30000 Da. Après migration, les gels ont été colorés 10 minutes par du bleu de Coomassie dans une solution de 40% méthanol, 10% d'acide acétique, 0,1% bleu de Coomassie R250, puis
15 décolorés dans une solution 10% méthanol, 10% acide acétique, 1% glycérol.

Quelque soit le système testé, la première observation est que la fréquence de transformation des
20 bactéries a été faible. Pour les bactéries qui ont pu être sélectionnées, le résultat des tests d'expression a été systématiquement négatif. Un exemple est donné sur la figure 4, panneaux A et B, avec les séries BL21(DE3)pLysS {[pT7-7], [pT7-7-TME1] ou [pT7-7-TME2]}.

25 Comme l'illustre la comparaison des courbes de croissance du panneau A de la figure 4, les inventeurs ont constaté avec les clones transformés avec pT7-7-TME1 ou pT7-7-TME2 et résistants sur milieu solide, que l'induction stoppe quasi-immédiatement la
30 croissance bactérienne, contrairement aux clones contenant le plasmide seul. De même, comme c'est

visible sur la figure 4(B), aucune bande de protéines migrant dans la région correspondant à la masse moléculaire des produits d'expression (~ 3-4000 Da) ou d'oligomères de ceux-ci ({1, 2, 3, ...}) x masse moléculaire) n'est en effet observable.

L'explication la plus probable à cette situation est que l'expression des domaines membranaires est très toxique pour la bactérie. La difficulté d'obtenir des transformants laisse supposer qu'une expression basale, même très faible, suffit à les tuer. Ceci montre aussi que le système pLysS n'est pas parfait pour prévenir cette expression basale. Parmi les bactéries qui résistent à l'étape de transformation, l'induction de l'expression des domaines hydrophobes devient immédiatement létale. Les systèmes utilisés permettent effectivement de protéger la bactérie hôte d'une expression basale, mais dès que celle-ci est induite, la toxicité est immédiate et les bactéries sont tuées.

20

Exemple 3 : Expression de séquences codant pour les protéines de fusion GST-TME1 et GST-TME2

Les vecteurs d'expression ont été construits comme décrit dans l'exemple 1, puis introduits dans les bactéries BL21(DE3)pLysS. Les bactéries BL21(DE3)pLysS ont été utilisées par soucis de comparaison avec les expériences précédentes car l'expression de la GST ou de ses chimères ne nécessite pas le système DE3-pLysS.

L'induction de l'expression a été effectuée avec de l'IPTG comme pour celle des domaines seuls. Les

caractéristiques des protéines produites sont résumées dans le tableau 2 suivant.

Tableau 2

Plasmide	Chimère, abréviation	Construction	Taille, aa	Masse, Da
pGEXKT	GST, G	<i>1</i> M-D ₂₃₉	239	27469
pGEXKT-T1	GST-TME1, GT1	<i>1</i> M-S ₂₃₃₋₃₄₇ M-A ₃₈₃	269	30506
pGEXKT-T2	GST-TME2, GT2	<i>1</i> M-S ₂₃₃₋₇₁₇ E-A ₇₄₆	263	30191

5

Les acides aminés (aa) sont indiqués avec le code à une lettre. La numérotation des séquences est effectuée par rapport aux protéines d'origine, GST et polyprotéine virale. Celle qui fait référence aux domaines membranaires est indiquée en italiques.

Les panneaux C et D de la figure 4 annexée montrent les résultats obtenus. Les courbes de croissance des bactéries transformées avec les différents plasmides montrent que l'expression des chimères GT1 et GT2 est toxique. Comme on peut le voir sur le gel d'électrophorèse de type Laemmli SDS-PAGE 14% [24], l'expression de TME1 fusionné à la GST s'accompagne de l'absence de bande migrant à la taille attendue de 30 kDa. Ceci laisse supposer qu'un très faible niveau d'expression de la chimère est suffisant pour tuer les bactéries. Par contre, la chimère GST-TME2 est cette fois visible sur le gel d'électrophorèse, dans la zone de masse moléculaire

attendue de 30 kDa. Le niveau d'expression reste toutefois limité.

La protéine produite n'est pas soluble malgré la présence de la GST dans la fusion. En effet, comme
5 le montre la figure 6 annexée, les essais de solubilisation, repliement et purification de la chimère GST-TME2 ont été un échec.

Pour obtenir les résultats représentés sur cette figure 6, les protéines GST et GST-TME2 ont été
10 exprimées comme décrit sur la figure 4 à partir de 150 ml de milieu de culture. Les bactéries ont ensuite été récoltées par centrifugation et suspendues KPO₄ 20 mM pH 7.7, NaCl 0.1M, EDTA 1 mM, NaN₃ 1 mM) pour avoir 100 DO/ml. Deux ml de chaque culture sont prélevés pour
15 sonication par pulses de 30 sec à 15% d'amplitude. Après sonication, Un prélèvement est effectué pour électrophorèse. Il correspond au puits « To » sur la figure 6 (correspondant au « total »).

Une première centrifugation à basse vitesse
20 (5000xg, 15 minutes) permet de séparer les bactéries non cassées et les corps d'inclusion des protéines solubles ou membranaires. Ces dernières se retrouvent dans le surnageant et un prélèvement est effectué. Il correspond au puits « Surn » sur la figure 6.

25 La fraction contenant la GST seule est ensuite traitée avec une résine d'affinité qui permet de fixer puis d'éluer spécifiquement cette protéine (puits « Af » du gel GST sur la figure 6).

La fraction contenant la protéine non soluble
30 GST-TME2 est traitée soit avec un détergent doux comme le triton X100 (TX100), en présence ou en absence de

NaCl, soit avec un détergent plus solubilisant mais plus destructurant comme le sarkosyl, avant d'être à nouveau diluée dans du TX100 et passée sur résine d'affinité.

5 Les résultats de la figure 6 montrent que la GST est présente dans la fraction soluble, à la différence de la fusion GST-TME2, ce qui indique que cette dernière est insoluble. Le surnageant contenant la GST est passé sur une résine d'agarose-GSH capable de fixer
10 la GST. Celle-ci est ensuite éluée avec un excès de GSH (puits indiqué « Af » du gel GST sur la figure 6).

Le culot contenant la fusion GST-TME2 n'est pas solubilisé en présence d'un détergent doux comme le TX100 (avec ou sans NaCl ajouté, puits « TX100 +/-
15 NaCl » du gel GST-TME2) mais il peut être solubilisé avec un détergent plus agressif comme le sarkosyl. Cependant, après dilution de la protéine ainsi solubilisée dans du TX100, un détergent doux qui doit favoriser son repliement, la protéine n'est pas retenue
20 sur la résine d'affinité, contrairement à la GST, ce qui suggère que la protéine de fusion ne peut être repliée.

Ces tests indiquent de façon claire que la protéine GST-TME2 est produite sous forme de corps
25 d'inclusion que l'on ne peut pas replier correctement.

Exemple 4 : Expression de vecteurs d'expression codant pour les protéines de fusion incluant un site Asp-Pro et un site GST

30 La construction des vecteurs a été réalisée comme décrit plus haut et pour les deux vecteurs codant

pour les protéines chimères GST-TME1 et GST-TME2, pour donner naissance aux vecteurs codant pour les protéines chimères GST-Asp-Pro-TME1 et GST-Asp-Pro-TME2. Elles sont résumées dans le tableau 3 ci-dessous.

5

Tableau 3

Plasmide	Chimère ; Abréviation fig. 4	Construction	Taille, aa	Masse, Da
pGEXKT- dp-T1	GST-DP-TME1 ; <i>G_{DP}T1</i>	₁ M-S ₂₃₃ dp- ₃₄₇ M- A ₃₈₃	271	30718
pGEXKT- dp-T2	GST-DP-TME2 ; <i>G_{DP}T2</i>	₁ M-S ₂₃₃ dp- ₇₁₇ E- A ₇₄₆	265	30403

Les acides aminés (aa) sont indiqués avec le code à une lettre. La numérotation des séquences est effectuée par rapport aux protéines d'origine, GST et polyprotéine virale. Celle qui fait référence aux domaines membranaires est indiquée en italiques.

Les vecteurs ont été testés comme décrit dans le paragraphe précédent. Les résultats obtenus sont illustrés sur les panneaux E et F de la figure 4 annexée.

Les courbes de croissance des bactéries transformées avec les différents plasmides montrent que l'expression des chimères *G_{dp}T1* et *G_{dp}T2* est nettement moins toxique que dans les cas précédents. Le panneau F montre que cette fois, TME1 est produite du fait de la présence du site de clivage DP. Son niveau d'expression, tel qu'on peut le voir dans le panneau F, est relativement moyen mais significatif. GST-DP-TME2

est nettement surproduit. Les deux protéines migrent dans leur zone de masse moléculaire attendue.

L'effet de l'addition du dipeptide DP est aussi significatif qu'inattendu : il amplifie l'expression des domaines et supprime leur toxicité. Cet effet
5 d'atténuation de la toxicité n'est pas connu pour le dipeptide DP, dont la seule propriété rapportée à ce jour est sa capacité à être coupé par l'acide formique. L'effet étant observé sur deux peptides différents et
10 tous deux initialement toxiques pour la bactérie, on peut donc raisonnablement penser que cette propriété peut s'étendre à d'autres peptides hydrophobes et toxiques.

Les inventeurs ont vérifié que le site peut
15 être effectivement clivé par l'acide formique : la coupure est lente et demande environ 7 jours à température ambiante.

Les essais d'expression à basse température (20°C) pendant toute une nuit de ces chimères ont
20 permis de mettre en évidence qu'elles sont produites sous forme native. En effet, il est possible de détecter une activité GST transférase dans la fraction membranaire des bactéries. De plus, cette activité est mesurée en solution lorsque les membranes sont
25 solubilisées en présence d'un détergent non ionique comme le β -D dodécylmaltoside, après centrifugation.

Exemple 5 : Expression de vecteurs d'expression codant pour les protéines de fusion incluant un site Asp-Pro et un site codant pour la thiorédoxine (TrX)

Les vecteurs d'expression pET32a-TrX, pET32a-TrX-dp-TME1 et pET32a-TrX-dp-TME2 ont été construits comme décrit ci-dessus et ont été ensuite introduits dans les bactéries BL21(DE3)pLysS. Les bactéries BL21(DE3)pLysS ont été utilisées par soucis de comparaison avec les expériences précédentes car l'expression de la GST ou de ses chimères ne nécessite pas le système DE3-pLysS. Les clones positifs ont été mis en culture et induits comme décrit ci-dessus.

L'induction de l'expression a été effectuée avec de l'IPTG comme pour celle des domaines seuls. Les caractéristiques des protéines produites sont résumées dans le tableau 4 suivant.

Tableau 4*

Plasmide	Chimère ; Abréviation fig. 4	Construction	Taille, aa	Masse, Da
pET32a	Thiorédoxine ; TrX	₁ M-C ₁₈₉	189	20397
pET32a- Gend-dp-T1	TrX-DP-TME1 ; T _{DP} T1	₁ M-S ₁₁₅ -PK- Gend-dp-T ₁	171	17796
pET32a- Gend-dp-T2	TrX-DP-TME2 ; T _{DP} T2	₁ M-S ₁₁₅ -PK- Gend-dp-T ₂	165	17481

* : T1 = TME1 et T2 = TME2

20

Les acides aminés (aa) sont indiqués avec le code à une lettre. La numérotation des séquences est effectuée par rapport aux protéines d'origine, GST et polyprotéine virale. Celle qui fait référence aux

domaines membranaires est indiquée en italiques.
« Gend » fait référence à la séquence C-terminale de la
GST provenant des constructions avec le plasmide
pGEXKT. Elle correspond à la séquence peptidique
5 primaire SDLSGGGGGLVPRGS. Les chimères thiorédoxine-
SDLSGGGGGLVPRGS-DP-(TME1 ou TME2) sont plus courtes que
la protéine codée dans le vecteur d'origine car
l'insertion est effectuée immédiatement après la
thiorédoxine.

10.

Sur la figure 5 :

A : la croissance bactérienne a été suivie en mesurant
l'augmentation de turbidité de chaque culture par
densité optique à 600 nm en fonction du temps.

15 **B :** les bactéries ont été prélevées comme indiqué pour
la figure 4. Elles ont ensuite été déposées sur gel
d'électrophorèse 14% d'acrylamide de type Laemmli
SDS-PAGE et traitées comme indiqué pour la figure 4.

20 Comme attendu, et comme le montrent les courbes
de croissance représentées sur la figure 5A annexée des
bactéries transformées avec les différents plasmides,
l'expression des chimères TrX-DP-TME1 et TrX-DP-TME2
suivant la présente invention n'est pas toxique. Le gel
25 d'électrophorèse représenté sur la figure 5B annexé de
type Laemmli SDS-PAGE 14% [24] montre que chaque
chimère est surproduite.

30 La présente invention permet donc de fabriquer
par recombinaison génétique des peptides hydrophobes
correspondant aux domaines membranaires des protéines

E1 et E2 de l'enveloppe du virus de l'hépatite C, dont l'expression a été reconnue comme létale dans les techniques de l'art antérieur. En outre, l'effet étant observé sur deux peptides vraiment différents et tous deux initialement toxiques pour la bactérie, indique que la présente invention concerne d'autres peptides hydrophobes et toxiques.

Exemple 6 : Effet du dipeptide DP sur la toxicité des domaines transmembranaires TME1 et TME2 exprimés sans protéine de fusion dans la bactérie

Cet exemple permet d'évaluer l'effet antitoxique du dipeptide DP inséré en absence de protéine de fusion GST ou TrX conformément à la revendication 1 annexée.

15

A) Matériels : Les plasmides pT7-7-pt_{TME1} et pT7-7-pt_{TME2} sont ceux qui sont décrits dans l'exemple 1. Les plasmides pT7-7-dp-pt_{TME1} et pT7-7-dp-pt_{TME2} ont été construits et clonés dans pT7-7 (SEQ IDn°33) comme décrits dans l'exemple 1 mais en utilisant les sites Nde I (5') et EcoR I (3') du plasmide. Les oligonucléotides amont (5') intègrent la séquence dp (gaccgcg) après la 1ère méthionine (atg). Les matrices utilisées pour générer chaque ADN ont été les plasmides pT7-7-pt_{TME1} et pT7-7-pt_{TME2}. Les séquences ont été vérifiées après clonage.

Les oligonucléotides utilisés sont les suivants :

i) Clonage de la séquence codant pour (M)DP-TME1 dans pT7-7:

OL19 (+) : 5'- CG CATATGGACCCGATCGCTGGTGCT -3'
(Nde I souligné) = (SEQ IDn°15 de la liste de séquences
annexée) ;

OL20 (-) : 5'-GAATTCCTAAGCGTCAACACCAGC-3' (EcoR
I souligné) = (SEQ IDn°16 de la liste de séquences
annexée).

ii) Clonage de la séquence codant pour (M)DP -
TME2 dans pT7-7:

OL28 (+) : 5'- CGCATATGGACCCGGAATACGTTGTTC-3'
(Nde I souligné) = (SEQ IDn°26 de la liste de séquences
annexée) ;

OL29 (-) : 5'-CAGAATTCCTAAGCTTCAGCCTGAGAG-3'
(EcoR I souligné) = (SEQ IDn°27 de la liste de
séquences annexée).

15

Les vecteurs d'expressions pT7-7-dp-pt_{TME1} et
pT7-7-dp-pt_{TME2} obtenus sont présentés sur la liste de
séquences annexée (SEQ IDn°44 et SEQ IDn°45).

20 **B) Légende des Figures 7A et B annexées :** la
souche bactérienne BL21(DE3)pLysS a été transformée
soit avec le plasmide seul soit avec les différentes
versions de pT7-7 intégrant les 4 constructions
exprimant TME1, M-DP-TME1 (figure 7A) ou TME2, M-DP-
25 TME2 (figure 7B). M représente la méthionine, elle est
présente en position N-terminale des peptides lorsque
les protéines toxiques sont fabriquées suivant la
présente invention avec le plasmide pT7-7.

La croissance des différents clones a été
30 comparée après induction par l'IPTG, suivant le
protocole identique à l'induction chimique décrite dans

l'exemple 2, et moyennée sur les valeurs de DO de 4 clones différents pour chaque construction.

C) Résultats:

5 Les figures 7A et 7B montrent que les bactéries ayant un plasmide exprimant des protéines TME1 et TME2 croissent moins vite après induction que la souche témoin qui est transformée avec le vecteur pT7-7 seul.

10 Ces résultats montrent que les souches transformées avec les plasmides exprimant les versions M-DP-TME1 (SEQ IDn°50) et M-DP-TME2 (SEQ IDn°51) selon l'invention croissent significativement mieux que celles qui expriment les TM sans DP. Ceci est vrai pour
15 TME1 et encore plus nettement pour TME2.

La conclusion est que l'insertion N-terminale de DP conformément à la présente invention contribue de manière surprenante à une diminution significative de la toxicité de l'expression des domaines membranaires,
20 notamment en absence d'une protéine de fusion soluble comme la GST ou la thiorédoxine.

Liste des références

- [1] Christendat D., Yee A., Dharamsi A., Kluger Y., Gerstein M., Arrowsmith C.H., and Edwards A.M., (2000),
5 *Prog. Biophys. Mol. Biol.* **73**, 339-345 ;
- [2] Hammarstrom M., Hellgren N., Van Den Berg S., Berglund H., and Hard T., (2002), *Protein Sci.* **11**, 313-321 ;
- [3] Falson P. (1992), *Biotechniques* **13**, 20-22 ;
- 10 [4] Falson P., Penin F., Divita G., Lavergne J.P., Di Pietro A., Goody R.S., and Gautheron D.C. (1993), *Biochemistry* **32**, 10387-10397 ;
- [5] Ciccagliione A.R., Marcantonio C., Costantino A., Equestre M., Geraci A. and Rapicetta M. (2000) *Virus*
15 *Genes* **21**, 223-226 ;
- [6] Sisk W.P., Bradley J.D., Kingsley D., and Patterson T.A. (1992) *Gene* **112**, 157-162 ;
- [7] Paulsen I.T., Sliwinski M.K., Nelissen B., Goffeau A., and Saier M.H. Jr. (1998) *FEBS Lett* **430**, 116-125 ;
- 20 [8] Decottignies A. and Goffeau A. (1997) *Nat Genet* **15**, 137-145 ;
- [9] Arechaga I., Miroux B., Karrasch S., Huijbregts R., de Kruijff B., Runswick M.J. and Walker J.E. (2000) *FEBS Lett* **482**, 215-219 ;

- [10] Miroux B. and Walker J.E. (1996) *J. Mol. Biol.* **260**, 289-298 ;
- [11] Mayo M.A., and Pringle C.R. (1998) *J. Gen Virol.* **79** (Pt4), 649-657 ;
- 5 [12] Op De Beeck A., Montserret R., Duvet S., Cocquerel L., Cacan R., Barberot B., Le Maire M., Penin F. and Dubuisson J. (2000) *J Biol Chem* **275**, 31428-31437 ;
- [13] Choo Q.L., Kuo G., Weiner A.J., Overby L.R., Bradley D.W., and Houghton M. (1989) *Science* **244**, 10 359-362 ;
- [14] Ciccaglione A.R., Marcantonio C., Costantino A., Equestre M., Geraci A. and Rapicetta M. (1998) *Virology* **250**, 1-8 ;
- [15] Ciccaglione A.R., Marcantonio C., Equestre M., Jones I.M. and Rapicetta M. (1998) *Virus Res* **55**, 157-165 ;
- [16] Op De Beeck A., Cocquerel L., and Dubuisson J. (2001) *J Gen Virol* **82**, 2589-2595 ;
- [17] Sharp P.M., Cowe E., Higgins D.G., Shields D.C., 20 Wolfe K.H., and Wright F. (1998) *Nucleic Acids Res* **16**, 8207-8211 ;
- [18] Mullis K.B., and Faloona F.A. (1987) *Methods Enzymol* **155**, 335-350 ;
- [19] Tabor S. and Richardson C.C. (1985) *Proc Natl Acad Sci USA* **82**, 1074-1078 ;
- 25

- [20] Guan K.L., and Dixon J.E. (1991) *Anal Biochem* **192**, 262-267 ;
- [21] Hakes D.J., and Dixon J.E. (1992) *Anal Biochem* **202**, 293-298 ;
- 5 [22] Tabor S. (1990) in *Current Protocols in Molecular Biology*, pp. 16.12.11-16.12.11, Greene Publishing and Wiley-Interscience, New York ;
- [23] Schagger H. and von Jagow G. (1987) *Anal Biochem* **166**, 368-379 ;
- 10 [24] Laemmli U.K. (1970) *Nature* **227**, 680-685.
- [25] Sambrook, Fritsch and Maniatis, *Molecular cloning, A laboratory manual*, second edition, Cold spring Harbor Laboratory Press, 1989.

15

20

25

REVENDICATIONS

1. Système d'expression caractérisé en ce qu'il comprend successivement, dans le sens 5'-3', une
5 séquence nucléotidique codant pour le dipeptide Asp-Pro et une séquence nucléotidique codant pour une protéine toxique membranaire ou un domaine d'une protéine toxique membranaire.
- 10 2. Système d'expression selon la revendication 1, dans lequel la protéine toxique est une protéine membranaire ou un domaine d'une protéine membranaire d'une enveloppe de virus.
- 15 3. Système d'expression selon la revendication 2, dans lequel le virus est choisi parmi le virus de l'hépatite C, le virus du sida, un virus pathogène pour l'homme, un virus pathogène pour un mammifère.
- 20 4. Système d'expression selon la revendication 1, dans lequel la protéine toxique est une protéine transmembranaire ou un domaine d'une protéine transmembranaire du virus de l'hépatite C.
- 25 5. Système d'expression selon la revendication 1, dans lequel la protéine toxique est une protéine de séquence IDn°1 ou IDn°2 de la liste de séquences annexée.
- 30 6. Système d'expression selon la revendication 1, dans lequel la séquence nucléotidique codant pour la

protéine toxique est choisie parmi la séquence IDn°3 et la séquence IDn°4 de la liste de séquences annexée.

7. Système d'expression selon la revendication
5 6, dans lequel la séquence nucléotidique codant pour le dipeptide Asp-Pro est gacccg.

8. Système d'expression selon la revendication
1, comprenant en outre, en amont de la séquence Asp-
10 Pro, une séquence nucléotidique codant pour une protéine soluble.

9. Système d'expression selon la revendication
8, dans lequel la protéine soluble est la glutathion S-
15 transférase ou la thiorédoxine.

10. Système d'expression selon la revendication
1, codant pour une protéine de fusion ayant une
séquence choisie dans le groupe constitué des séquences
20 IDn°46, IDn°47, IDn°48, IDn°49, IDn°50 et IDn°51 de la liste de séquences annexée.

11. Système d'expression selon la revendication
8, ledit système ayant une séquence choisie dans le
25 groupe constitué des séquences IDn°34, IDn°35, IDn°36, IDn°37, IDn°38, et IDn°39 de la liste de séquences annexée.

12. Vecteur d'expression bactérien comprenant
30 un système d'expression selon l'une quelconque des revendications 1 à 11 cloné dans un plasmide.

13. Vecteur d'expression bactérien comportant un système d'expression selon l'une quelconque des revendications 1 à 9 et la séquence oligonucléotidique du plasmide pT7-7.

14. Vecteur d'expression bactérien constitué de la séquence IDn°44 ou IDn°45 de la liste de séquences annexée.

15. Vecteur d'expression bactérien comportant un système d'expression selon l'une quelconque des revendications 1 à 7 et la séquence oligonucléotidique d'un plasmide choisi parmi pGEXKT, pET32a.

16. Vecteur d'expression bactérien selon la revendication 15 constitué d'une séquence choisie dans le groupe constitué des séquences IDn°40, IDn°41, IDn°42 et IDn°43 de la liste de séquences annexées.

17. Cellule procaryote transformée par un vecteur d'expression selon l'une quelconque des revendications 12 à 16.

18. Cellule procaryote *E. coli* selon la revendication 17.

19. Procédé de fabrication d'une protéine toxique par recombinaison génétique comprenant les étapes suivantes :

- transformer une cellule hôte avec un système d'expression selon la revendication 1 ou avec un vecteur d'expression selon la revendication 12,

- cultiver la cellule hôte transformée dans des conditions de culture telles qu'elle fabrique une protéine de fusion comprenant le dipeptide Asp-Pro suivi de la séquence peptidique de la protéine toxique à partir dudit vecteur d'expression, et

- isoler ladite protéine de fusion.

10

20. Procédé selon la revendication 19, comprenant en outre une étape consistant à cliver ladite protéine de fusion pour récupérer la protéine toxique.

15

21. Procédé selon la revendication 20, dans lequel l'étape consistant à cliver ladite protéine de fusion pour récupérer la protéine toxique est réalisée en faisant agir de l'acide formique sur la protéine de fusion.

20

22. Procédé selon la revendication 19, dans lequel la cellule hôte est *E. coli*.

25

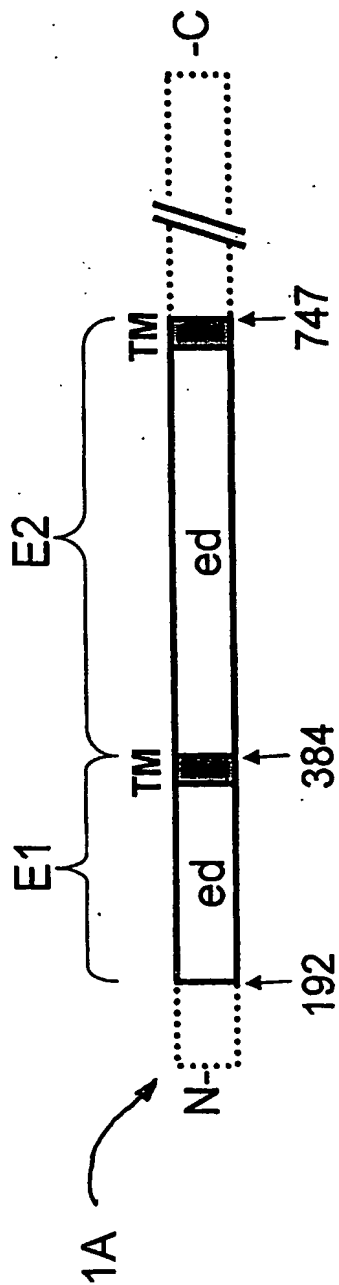
23. Procédé selon la revendication 19, dans lequel le système d'expression code pour une protéine de fusion ayant une séquence choisie dans le groupe constitué des séquences IDn°46, IDn°47, IDn°48, IDn°49, IDn°50 et IDn°51 de la liste de séquences annexée.

30

24. Procédé selon la revendication 19, dans lequel le système d'expression a une séquence choisie dans le groupe constitué des séquences IDn°34, IDn°35, IDn°36, IDn°37, IDn°38, et IDn°39 de la liste de
5 séquences annexée.

25. Procédé selon la revendication 19, dans lequel le vecteur d'expression est constitué d'une séquence choisie dans le groupe constitué des séquences
10 IDn°40, IDn°41, IDn°42, IDn°43, IDn°44 et IDn°45 de la liste de séquences annexées.

26. Protéine de fusion ayant une séquence peptidique choisie dans le groupe constitué des
15 séquences IDn°46, IDn°47 IDn°48, IDn°49, IDn°50 et IDn°51 de la liste de séquences annexée.



TME1: 347-MIAGAHWGVLGIA⁻YFSMVGNWAKVLVLLLEAGVDA-383

*** * * * * *** * * * * * * * * * *

TME2: 717-EYVLLFLLADARVCSCIWMMLLISQAEA-746

***** * * * * ***** * *

FIG. 1

2 / 9

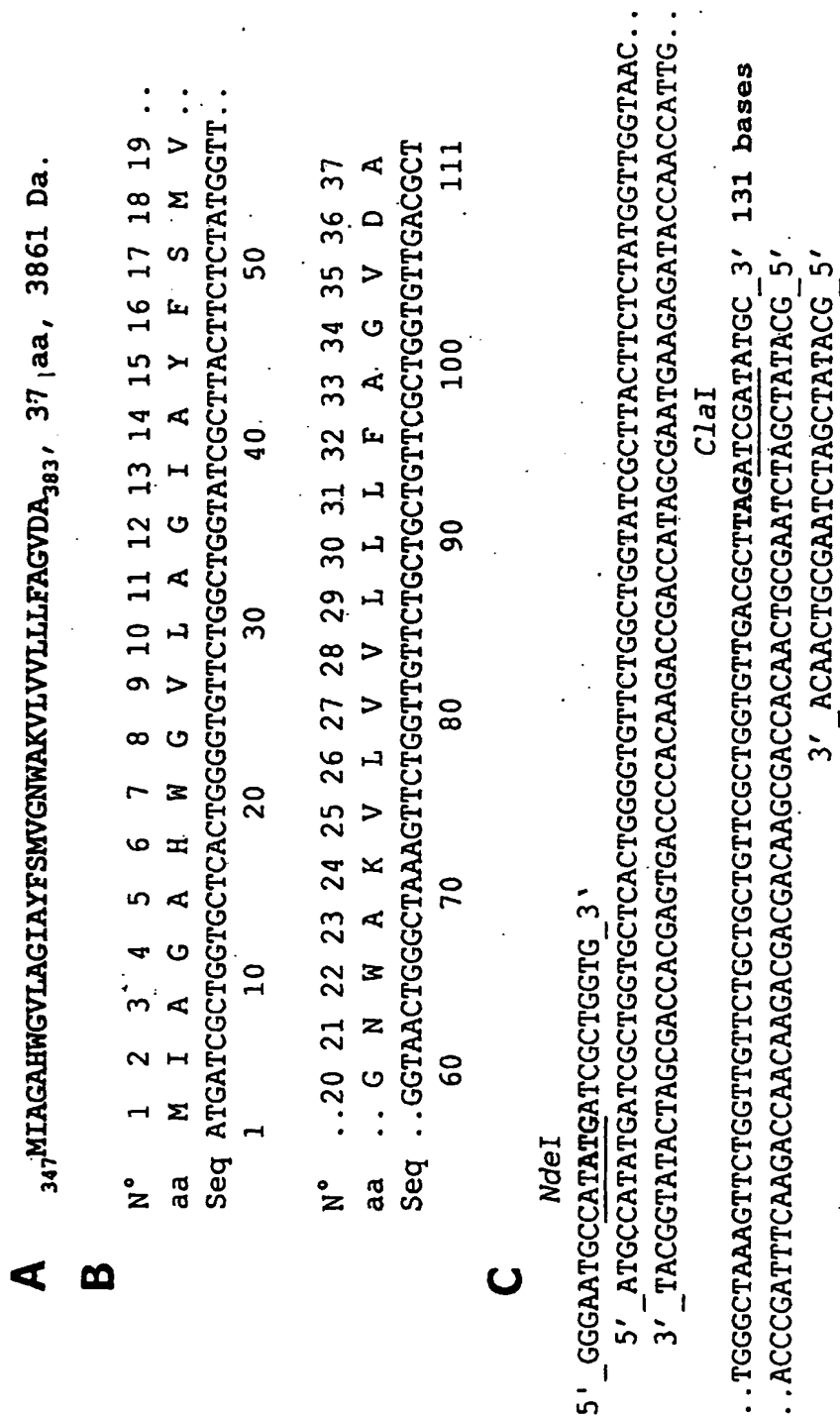


FIG. 2

3 / 9

D

OL11(+): 5'atgccatgatcgctggtgctcactggggttctggtggtatcgcttactctctatggttgtaactggg
OL12(-): 5'gcataatgatctaaagctcaacaccagcgaacagcagacagacagaaccagaacttttagcccgagttaccaaccatagagaa

Clonage dans pT7-7:

OL13(+): 5'gggaatgccatgatgatcgctggtg
OL14(-): 5'gcataatgatctaaagctcaaca

Clonage dans pGEXKT:

OL15(+): 5'ggatccatggaatacgtgttc (sans site DP)
OL17(+): 5'ggatcccgaaccgatggaatacgtgttc (avec site DP)
OL16(-): 5'gaattcctaagcttcagcctgag

Clonage dans pET32a:

OL18(+): 5'gtgatatctgatctgtctggtggtgt (s'hybride au segment 915-932 de pGEXKT)
OL16(-): 5'gaattcctaagcttcagcctgag

Clonage dans pT7-7 de (M)DP-TME1 :

OL19 (+) : 5'- CGCATAATGGACCCGATCGCTGGTGCT -3' (Nde I souligné)
OL20 (-) : 5'-GAATTCCTAAGCGTCAACACCAGC-3' (EcoR I souligné)

FIG. 2

4 / 9

A (M)₇₁₇ EYVLLFLLADARVCSCIWMMLLISQAEA₇₄₆ 31 aa, 3546 Da.

B

N°	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	..
Aa	M	E	Y	V	V	L	L	F	L	L	L	A	D	A	R	V	..
Seq	ATGGAATACGTTGTTCTGCTGTTCCCTGCTGCTGGCTGACGCTCGTGT..																
	1	10	20	30	40												

N°	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
Aa	..	C	S	C	L	W	M	M	L	L	I	S	Q	A	E	A
Seq	..TGCTCTTGCCCTGTGGATGATGCTGCTGCTGATCTCTCAGGCTGAAGCT															
	50	60	70	80	90	93										

C

NdeI

5'-CATATGGAATACGTTGTTTC_3'

(+)5'-CATATGGAATACGTTGTTCTGCTGTTCCCTGCTGCTGGCTGACGCTCGTGT..(-)3'-GTATACCTTATGCAACAAGACGACGACAAGGACGACGACCGACTGCGAGCACAA..

Hind III

..TGCTCTTGCCCTGTGGATGATGCTGCTGATCTCTCAGGCTGAAGCTTAAGCTT-3' 117 bp

..ACGAGAACGGACACCTACTACGACGACTAGAGAGTCCGACTTCGAAATTCGAA-5'

3'-GACTAGAGAGTCCGACTTCGAAATTCGAA-5'

FIG. 3

OL21(+): 5'catatggaatacgttctgtctgttctgtctgctggtgacgtcggtttgtctcttgcttggtgat
OL22(-): 5'aagcttaagcttcagcctgagagatcagcagcatcatccacaggcaagagcaaacac

Clonage dans pT7-7:

OL23(+): 5'catatggaatacgttgctc

OL24(-): 5'aagcttaagcttcagcctgagatcag

Clonage dans pGEXKT:

OL25(+): 5'qqatccgaatacgttggtc (sans site DP)

OL25(+): 5'ggatccgacccggaatacgtgttc
OL27(+): 5'ggatccgacccggaatacgtgttc (avec site DP)

OL26(-): 5'gaattctaagcttcagcctgagagatcag

Clonage dans pET32a:

Clonage dans pET3a.
 Q118(+): 5'atgaatctgaatctatctgaatgaagt (s'hybride au segment 915-932 de pGEXKT)

OL26(-): 5'gaattctaagcttcagcctgagagatcag

Clonage dans pT7-7 de (M)DP-TME2 :

Clivage dans p17-7 de (1/5) site :

0128 (+) : 5'- CGCATATGGACCCGGAATACGTTGTC-3' (Nde I souligné)

OL29 (-) : 5'-CAGAATTCCTAGCTTCAGCCTGAGAG-3' (EcoR I souligné)

3
G
F

6 / 9

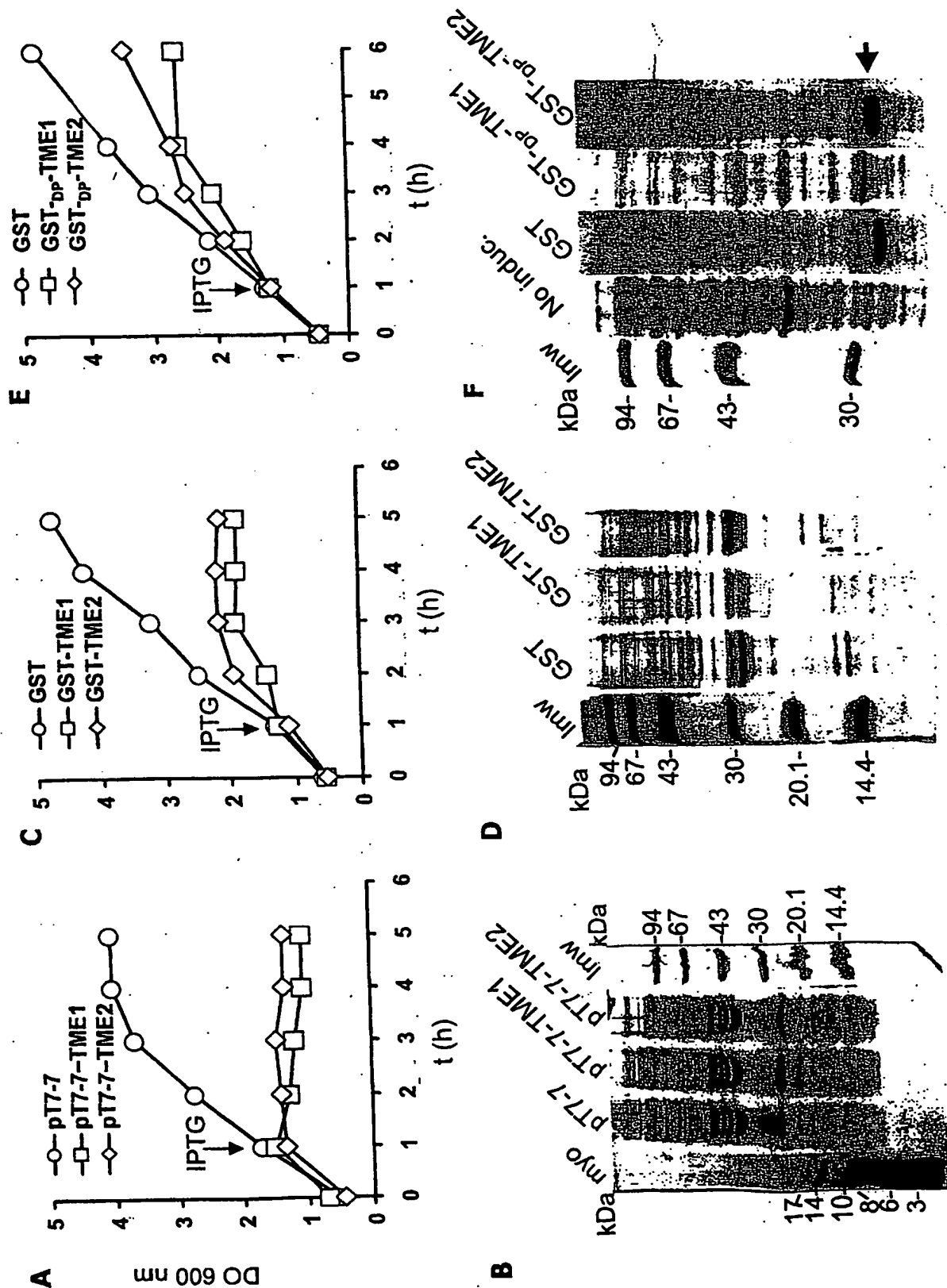


FIG. 4

7 / 9

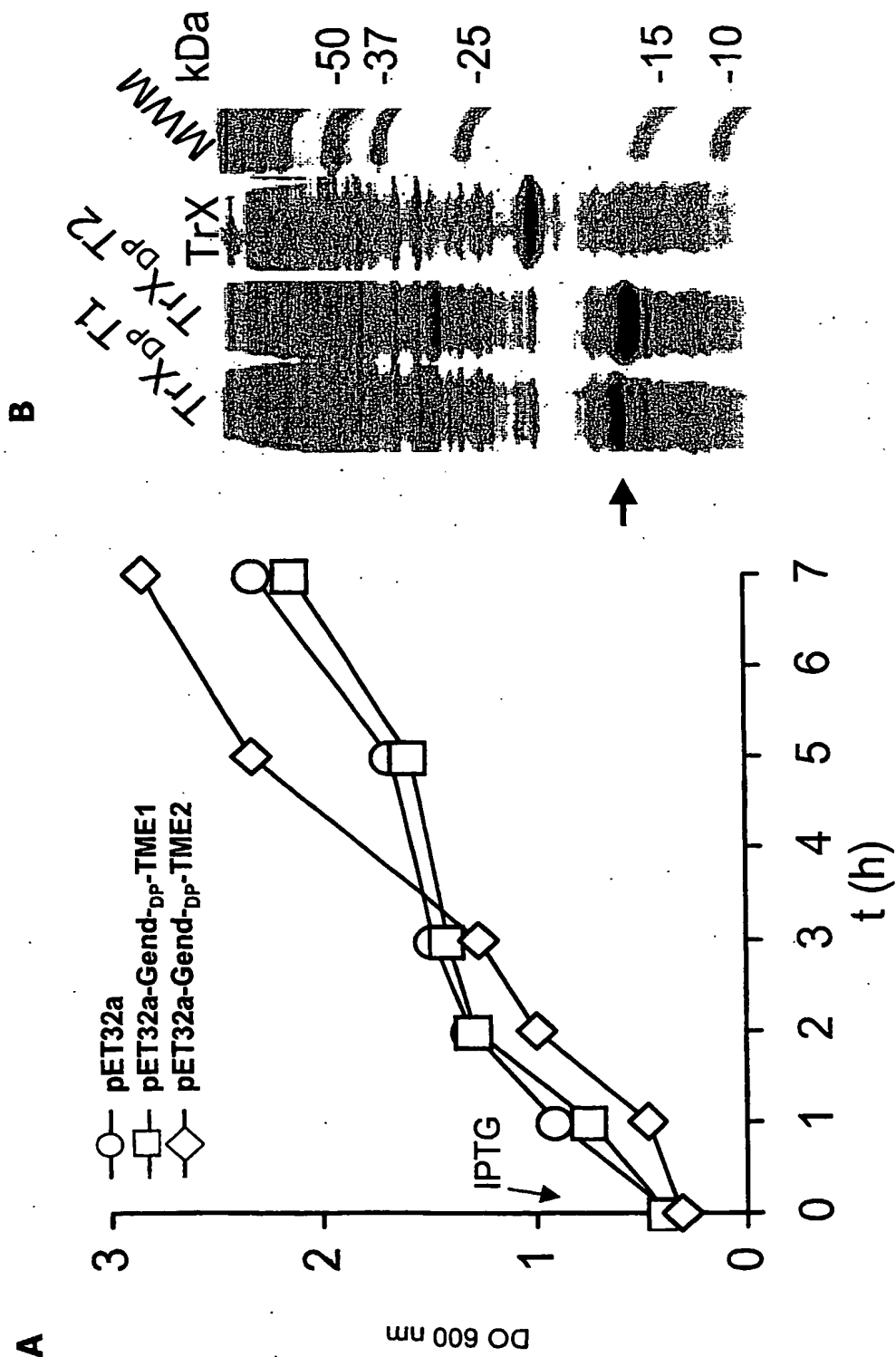


FIG. 5

GST :

MSPIIGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMARIYI
ADKHNMLGGCPKERAIEISMLEGAVLDIRYGVSRISYKDFETLKVDLFSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPOIDKYLKSSKYIAWPLQGWQATFGGDDHPPKSDLSGGG
GGLVPRGS/PGIHRD

GST-TME2 :

MSPIIGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMARIYI
DKHNMLGGCPKERAIEISMLEGAVLDIRYGVSRISYKDFETLKVDLFSKLPEMLKMFEDRLCHKTYLNGDHVTHPD
FMYLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPOIDKYLKSSKYIAWPLQGWQATFGGDDHPPKSDLSGGGGG
VPRGS/EYVLLFLLADARVCCLMMMLISQAEA

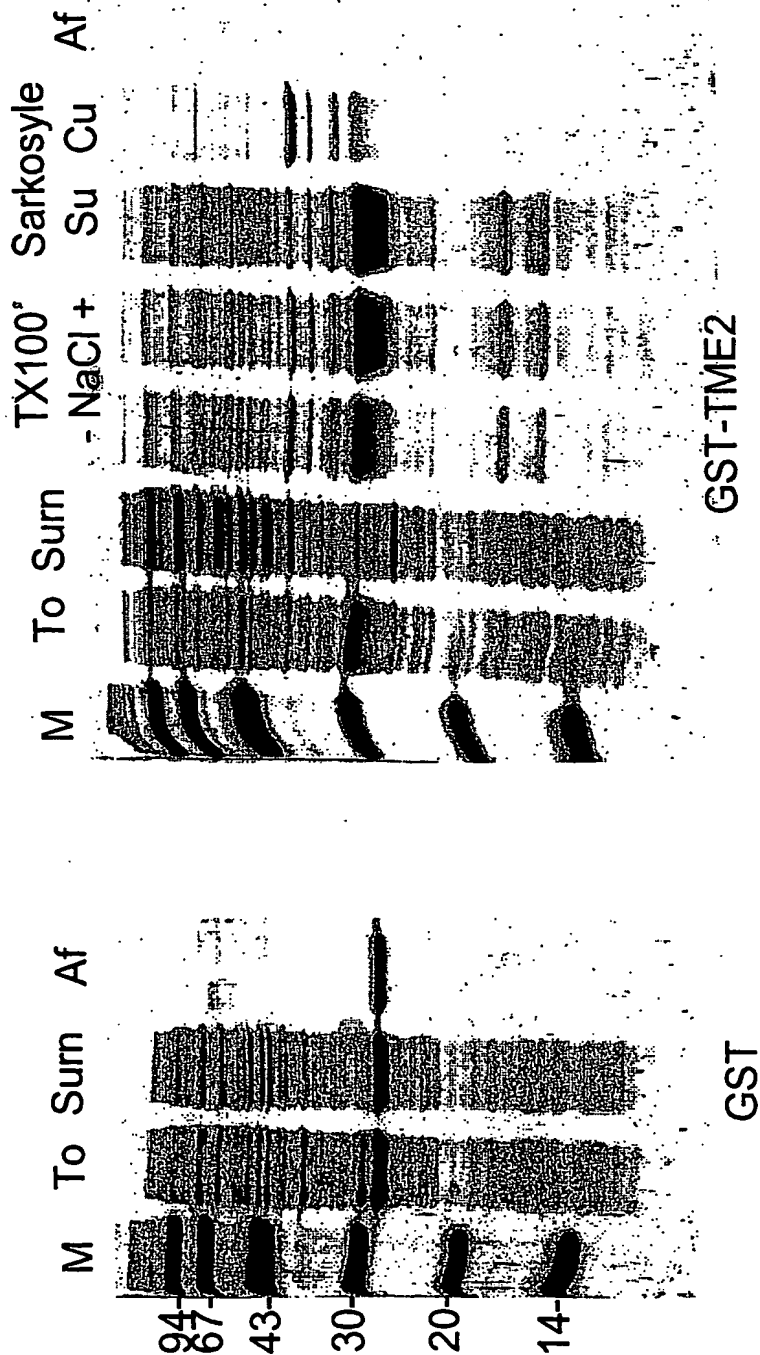


FIG. 6

9 / 9

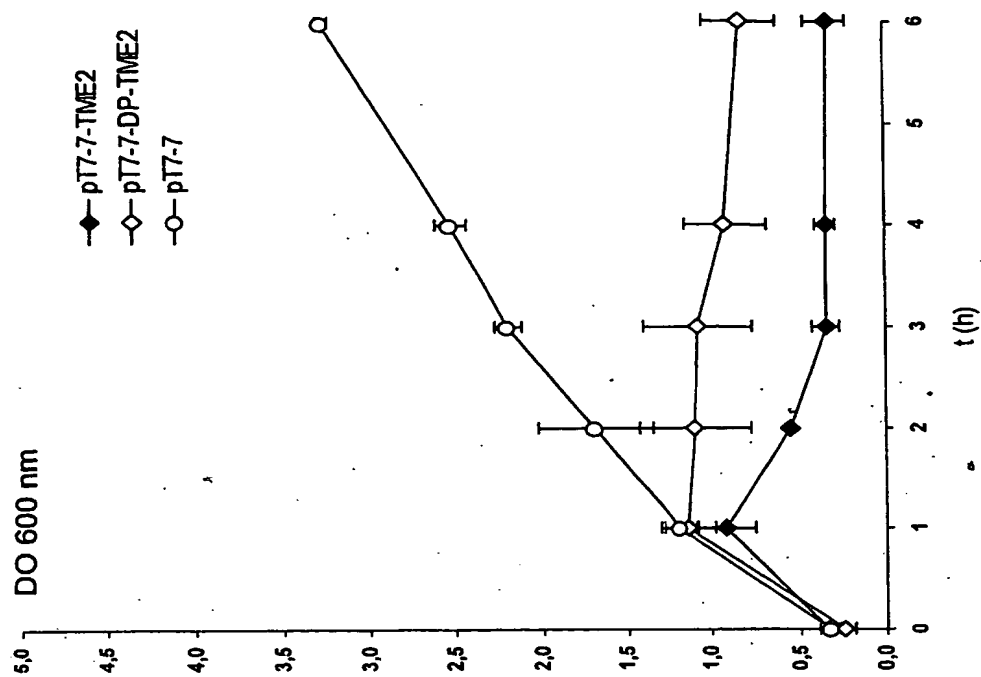


FIG. 7A

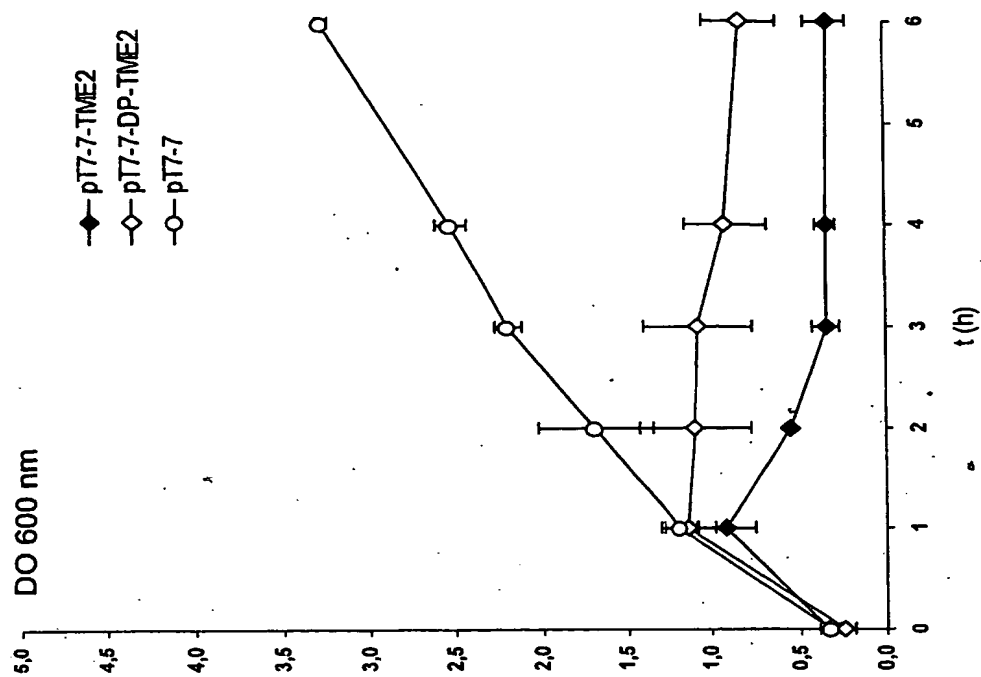


FIG. 7B

LISTE DE SEQUENCES

<110> CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

<120> SYSTEMES D'EXPRESSION DE PROTEINES TOXIQUES, VECTEURS
ET PROCEDE DE FABRICATION DE PROTEINES TOXIQUES

<130> B14143 EE

<140>

<141>

<150> FR N°02 11676

<151> 2002-09-20

<160> 53

<170> PatentIn Ver. 2.1

<210> 1

<211> 37

<212> PRT

<213> Hepatitis C virus

<400> 1

Met Ile Ala Gly Ala His Trp Gly Val Leu Ala Gly Ile Ala Tyr Phe
1 5 10 15Ser Met Val Gly Asn Trp Ala Lys Val Leu Val Val Leu Leu Phe
20 25 30Ala Gly Val Asp Ala
35

<210> 2

<211> 31

<212> PRT

<213> Hepatitis C virus

<400> 2

Met Glu Tyr Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala Arg Val
1 5 10 15Cys Ser Cys Leu Trp Met Met Leu Leu Ile Ser Gln Ala Glu Ala
20 25 30

<210> 3

<211> 111

<212> ADN

<213> Hepatitis C virus

<400> 3

atgatcgctg gtgctcaactg ggggtgttctg gctgggtatcg cttacttctc tatgggttgg 60
aactgggcta aagttctgggt tgttctgctg ctgttcgctg gtgttgacgc t 111

<210> 4

<211> 93

<212> ADN

<213> Hepatitis C virus

<400> 4
atggaatacgt ttgttctgct gttcttctgt ctggctgacg ctctgtgttg ctcttgcttg 60
tggatgatgc tgctgatctc tcaggctgaa gct 93

<210> 5
<211> 24
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:
oligonucléotide (+) d'insertion dans pT7-7

<400> 5
gggaatgccca tatgatcgct ggtg 24

<210> 6
<211> 23
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:
oligonucléotide (-) d'insertion dans pT7-7

<400> 6
gcatatcgat ctaagcgtca aca 23

<210> 7
<211> 131
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle: ADN sens
codant de TME1 + site cla I en 3'

<400> 7
atgccatattg atcgctgggtg ctcaactgggg tggtctgggt ggtatcgctt acttctctat 60
gggttgtaac tgggctaaag ttctgggttg tctgctgctg ttcgctgggtg ttgacgctta 120
gatcgatatg c 131

<210> 8
<211> 131
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle: ADN sens
anticodant + site cla I en 5'

<400> 8
gcatatcgat ctaagcgtca acaccagcga acagcagcag aacaaccaga actttagccc 60
agttaccaac catagagaag taagcgatac cagccagaac accccagtga gcaccagcga 120
tcatatggca t 131

<210> 9
<211> 74
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:
oligonucléotide sens codant de pour la synthèse de
TME1

<400> 9
atgccatatg atcgctgggtg ctccactgggg tgttctggct ggtatcgctt acttctctat 60
ggttggtaac tggg 74

<210> 10
<211> 79
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:
oligonucléotide sens anticodant pour la synthèse
de TME1

<400> 10
gcatatcgat ctaagcgtca acaccagcga acagcagcag aacaaccaga actttagccc 60
agttaccaac catagagaa 79

<210> 11
<211> 22
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:
oligonucléotide (+) d'insertion dans pGEXKT sans
site dp

<400> 11
ggatccatgg aatacgttgt tc 22

<210> 12
<211> 28
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:
oligonucléotide (+) d'insertion dans pGEXKT avec
site dp

<400> 12
ggatccgacc cgatggaata cgttgttc 28

<210> 13

<211> 23
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:
oligonucléotide (-) d'insertion dans pGEXKT

<400> 13
gaattcctaa gcttcagcct gag

23

<210> 14
<211> 27
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:
oligonucléotide (+) de transfert dans pET32a

<400> 14
gtgatatctg atctgtctgg tgggtgt

27

<210> 15
<211> 26
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:
oligonucléotide (+) d'insertion dans pT7-7

<400> 15
cgcatatgga cccgatcgct ggtgct

26

<210> 16
<211> 24
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:
oligonucléotide (-) d'insertion dans pT7-7

<400> 16
gaattcctaa gcgtcaacac cagc

24

<210> 17
<211> 19
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:
oligonucléotide (+) d'insertion dans pT7-7

<400> 17

catatggaat acgttggttc

19

<210> 18
 <211> 28
 <212> ADN
 <213> Séquence artificielle

<220>
 <223> Description de la séquence artificielle:
 oligonucléotide (-) d'insertion dans pT7-7

<400> 18
 aagcttaagc ttcagcctga gagatcag

28

<210> 19
 <211> 103
 <212> ADN
 <213> Séquence artificielle

<220>
 <223> Description de la séquence artificielle: ADN sens
 codant pour TME2 + site Nde I en 5' et Hind III en
 3'

<400> 19
 catatggaat acgttggttct gctgttcctg ctgctggctg acgctcgtgt ttgctcttgc 60
 ctgtggatga tgctgctgat ctctcaggct gaagcttaag ctt 103

<210> 20
 <211> 103
 <212> ADN
 <213> Séquence artificielle

<220>
 <223> Description de la séquence artificielle: ADN sens
 anticodant de TME2 + site Nde I en 3' et Hind III
 en 5'

<400> 20
 aagcttaagc ttcagcctga gagatcagca gcacatcca caggcaagag caaacacgag 60
 cgtcagccag cagcaggaac agcagaacaa cgtattccat atg 103

<210> 21
 <211> 68
 <212> ADN
 <213> Séquence artificielle

<220>
 <223> Description de la séquence artificielle:
 oligonucléotide sens codant (+) pour la synthèse
 de TME2

<400> 21
 catatggaat acgttggttct gctgttcctg ctgctggctg acgctcgtgt ttgctcttgc 60
 ctgtggat 68

<210> 22
<211> 57
<212> ADN
<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:
oligonucléotide sens codant (-) pour la synthèse
de TME2

<400> 22
aagcttaagc ttcagcctga gagatcagca gcatcatcca caggcaagac gaaacac 57

<210> 23
<211> 19
<212> ADN
<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:
oligonucléotide (+) d'insertion dans pGEXKT sans
site dp

<400> 23
ggatccgaat acgttggtc 19

<210> 24
<211> 25
<212> ADN
<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:
oligonucléotide (+) d'insertion dans pGEXKT avec
site dp

<400> 24
ggatccgacc ccgaatacgt tggtc 25

<210> 25
<211> 30
<212> ADN
<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:
oligonucléotide (-) d'insertion dans pGEXKT avec
site dp

<400> 25
gaattcttaa gcttcagcct gagagatcag 30

<210> 26
<211> 27
<212> ADN
<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:
oligonucléotide (+) d'insertion dans pT7-7

<400> 26

cgcatatgga cccggaatac gttgttc

27

<210> 27

<211> 27

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:
oligonucléotide (-) d'insertion dans pT7-7

<400> 27

cagaattcct aagcttcagc ctgagag

27

<210> 28

<211> 15

<212> PRT

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: fin de la
GST suivie du site thrombine

<400> 28

Ser	Asp	Leu	Ser	Gly	Gly	Gly	Gly	Gly	Leu	Val	Pro	Arg	Gly	Ser
1				5					10					15

<210> 29

<211> 717

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: ADN
codant pour la protéine GST dans le vecteur pGEXKT

<400> 29

atgtccccta	tactagggtta	ttggaaaatt	aagggccttg	tgcaaccac	tcgacttctt	60
ttggaatata	ttgaagaaaa	atatgaagag	catttgtatg	agcgcgatga	agggtgataaa	120
tggcgaaaca	aaaagtgttg	attgggtttg	gagtttccca	atcttcctta	ttatattgat	180
ggtgatgtta	aattaacaca	gtctatggcc	atcatagctt	atatagctga	caagcacaac	240
atgttgggtg	gttgtccaaa	agagcgtgca	gagatttcaa	tgcttgaagg	agcgggtttg	300
gatattagat	acggtgtttc	gagaattgca	tatagtaaag	actttgaaac	tctcaaagtt	360
gattttctta	gcaagctacc	tgaaatgctg	aaaatgttcg	aagatcgttt	atgtcataaa	420
acatatattaa	atggtgatca	tgtaacccat	cctgacttca	tgttgtatga	cgctcttgat	480
gttggttttat	acatggaccc	aatgtgcctg	gatgcgttcc	caaaattagt	ttgtttttaa	540
aaacgtattg	aagctatccc	acaaattgat	aagtacttga	aatccagcaa	gtatatagca	600
tggcctttgc	agggctggca	agccacgttt	ggtgggtggc	accatcctcc	aaaatcggat	660
ctgtctgggtg	gtgggtgggtg	tctgggttccg	cgtggatccc	cggaattca	tcgtgac	717

<210> 30

<211> 327

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: ADN
codant pour la thiorédoxine dans le vecteur
pET32a+

<400> 30

```
atgagcgata aaattattca cctgactgac gacagttttg acacggatgt actcaaagcg 60
gacggggcga tcctcgtcga tttctgggca gagggtgtcg gtccgtgcaa aatgatcgcc 120
ccgattcttg atgaaatcgc tgacgaatat cagggcaaac tgaccgttgc aaaactgaac 180
atcgatcaaa accctggcac tgcgccgaaa tatggcatcc gtggatatcc gactctgctg 240
ctgttcaaaa acggtgaagt ggcggcaacc aaagtgggtg cactgtctaa aggtcagttg 300
aaagagttcc tcgacgctaa cctggcc 327
```

<210> 31

<211> 4969

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: plasmide
d'expression pGEXKT

<400> 31

```
acgttatcga ctgcacgggtg caccaatgct tctggcgctca ggcagccatc ggaagctgtg 60
gtatggctgt gcaggctcgt aatcactgca taattcgtgt cgctcaaggc gcactcccgt 120
tctggataat gttttttgct cgcacatcat aacgggttctg gcaaatattc tgaaatgagc 180
tggtgacaat taatcatcgg ctctgataat gtgtggaatt gtgagcggat aacaatttca 240
cacaggaaac agtattcatg tcccctatac taggtttattg gaaaattaag ggccttgtgc 300
aaccactcgt acttcttttg gaatatcttg aagaaaaata tgaagagcat ttgtatgagc 360
gcgatgaagg tgataaatgg cgaaacaaaa agtttgaatt gggtttggag tttcccaatc 420
ttccttatta tattgatggg gatgttaaat taacacagtc tatggccatc atacgttata 480
tagctgacaa gcacaacatg ttgggtgggt gtccaaaaga gcgtgcagag atttcaatgc 540
ttgaaggagc ggttttggat attagatacg gtgtttcgag aattgcatat agtaaagact 600
ttgaaactct caaagttgat tttcttagca agctacctga aatgctgaaa atgttcgaag 660
atcgtttatg tcataaaaca tatttaaatg gtgatcatgt aacccatcct gacttcatgt 720
tgtatgacgc tcttgatggt gttttataca tggacccaat gtgcctggat gcgttcccaa 780
aattagtttg ttttaaaaaa cgtattgaag ctatcccaca aattgataag tacttgaaat 840
ccagcaagta tatagcatgg cctttgcagg gctggcaagc cacgtttggt ggtggcgacc 900
atcctccaaa atcgatctg tctgggtggg gtgggtggtc ggttcccgct ggatccccgg 960
gaattcatcg tgactgactg acgatctgcc tcgcgcgttt cggtgatgac ggtgaaaacc 1020
tctgacacat gcagctcccc gagacgggtc cagcttctct gtaagcggat gccgggagca 1080
gacaagcccg tcagggcgctg tcagcgggtg ttggcgggtg tcggggcgca gccatgaccc 1140
agtcacgtag cgatagcggg gtgtataatt cttgaagacg aaagggcctc gtgatacgcc 1200
tatttttata ggttaatgtc atgataataa tggtttctta gacgtcaggt ggcacttttc 1260
ggggaaatgt gcgcggaacc cctatttgtt tatttttcta aatacattca aatatgtatc 1320
cgctcatgag acaataaacc tgataaatgc ttcaataata ttgaaaaagg aagagtatga 1380
gtattcaaca tttccgtgtc gcccttattc ccttttttgc ggcattttgc cttcctgttt 1440
ttgctcaccg agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 1500
tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagtttt cgccccgaag 1560
aacgttttcc aatgatgagc acttttaaaag ttctgctatg tggcgcggta ttatcccggtg 1620
ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttgggtg 1680
agtaactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca 1740
gtgctgccat gataacactg cggccaactt acttctgaca acgatcggag 1800
gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc 1860
gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 1920
cagcaatggc aacaacggtg cgcaactat taactggcga actacttact ctacttccc 1980
ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg 2040
```

```

cccttcgggc tggctgggtt attgctgata aatctggagc cggtgagcgt gggctctcgc 2100
gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga 2160
cggggagtcg ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 2220
tgattaagca ttggtaactg tcagaccaag tttactcata tatacttttag attgatttaa 2280
aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcattgacca 2340
aaatccctta acgtgagttt tcgttcact gagcgtcaga ccccgtagaa aagatcaaag 2400
gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac 2460
cgctaccagc ggtgggtttgt ttgccggatc aagagctacc aactottttt ccgaaggtaa 2520
ctggcttcag cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc 2580
acacttcaa gaactctgta gcaccgccta catacctcgc tctgctaato ctgttaccag 2640
agctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 2700
cggataaggc gcagcggctg ggctgaacgg ggggttcgtg cacacagccc agcttgaggc 2760
gaacgcctta caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc 2820
ccgaaggagc aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca 2880
aggggagct tccaggggga aacgcctggg atctttatag tctgtcggg tttcgccacc 2940
ctgacttga gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg 3000
cagcaacgc ggccttttta cggttcctgg ccttttgctg gccttttgct cacatgttct 3060
cctcgcgtt atccctgat tctgtggata accgtattac cgcctttgag tgagctgata 3120
agctcgcgc cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc 3180
agctgatgcg gtattttctc cttacgcata tctgcggtat ttcacaccgc ataaattccg 3240
acaccatcga atggtgcaaa acctttcgcg gtatggcatg atagcgcccg gaagagagtc 3300
aattcagggt ggtgaatgtg aaaccagtaa cgttatacga tgtcgagag tatgccgggtg 3360
tctcttatca gaccgtttcc cgcgtgggta accaggccag ccacgtttct gcgaaaacgc 3420
gggaaaaagt ggaagcggcg atggcggagc tgaattacat tcccaaccgc gtggcacaac 3480
aactggcggg caaacagtcg ttgctgattg gccttgccac ctccagtctg gccctgcacg 3540
cgccgtcgca aattgtcgcg gcgattaaat ctcgcgcgca tcaactgggt gccagcgtgg 3600
tggtgtcgat ggtagaacga agcggcgtcg aagcctgtaa agcggcgggt cacaatcttc 3660
tcgcgcaacg cgtcagtggt ctgatcatta actatccgct ggatgaccag gatgccattg 3720
ctgtggaagc tgcctgcact aatgttccgg cgttatttct tgatgtctct gaccagacac 3780
ccatcaacag tattattttc tcccatgaag acggtacgcg actgggcgtg gagcatctgg 3840
tcgcattggg tcaccagcaa atcgcgctgt tagcgggccc attagttct gtctcggcgc 3900
gtctgcgtct ggctggctgg cataaatatc tcaactgcaa tcaaattcag ccgatagcgg 3960
aacgggaagg cgactggagt gccatgtccg gttttcaaca aaccatgcaa atgctgaatg 4020
agggcatcgt tccactgcg atgctgggtt ccaacgatca gatggcgtg ggcgcaatgc 4080
gcgccattac cgagtcgggg ctgcgcgttg gtgcggatat ctcggtagtg ggatacgacg 4140
ataccgaaga cagctcatgt tatatccgcg cgttaaccac catcaaacag gattttcgcc 4200
tgctggggca aaccagcgtg gacgcgttgc tgcaactctc tcagggccag gcggtgaagg 4260
gcaatcagct gttgcccgtc tcaactggta aaagaaaaac caccctggcg cccaatacgc 4320
aaaccgcctc tccccgcgcg ttggccgatt cattaatgca gctggcacga caggtttccc 4380
gactggaaag cgggcagtgga gcgcaacgca attaatgtga gttagctcac tcattaggca 4440
ccccaggctt tacactttat gcttccggct cgtatgttgt gtggaattgt gagcggataa 4500
caatttcaca caggaaacag ctatgaccat gattacggat tcaactggcg tegttttaca 4560
acgtcgtgac tgggaaaacc ctggcgttac ccaacttaat cgccttgca ccatcccc 4620
tttcgccagc tggcgtaata gcgaagaggc ccgcaccgat cgccttccc aacagttgcg 4680
cagcctgaat ggcgaatggc gctttgcctg gtttccggca ccagaagcgg tgccggaaag 4740
ctggctggag tgcgatcttc ctgaggccga tactgtcgtc gtccctcaa actggcagat 4800
gcacggttac gatgcgcca tctacaccaa cgtaacctat ccattacgg tcaatccgcc 4860
gtttgttccc acggagaatc cgacgggttg ttactcgctc acatttaatg ttgatgaaag 4920
ctggctacag gaaggccaga cgcgaattat ttttgatggc gttggaatt 4969

```

<210> 32

<211> 11800

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: plasmide d'expression pET32a+

<400> 32

atccggatat	agttcctcct	ttcagcaaaa	aacccctcaa	gaccggttta	gaggccccaa	60
ggggttatgc	tagttattgc	tcagcgggtg	cagcagccaa	taggcctata	tcaaggagga	120
aagtcgtttt	ttggggagtt	ctgggcaaat	ctccggggtt	ccccaatagc	atcaataacg	180
agtcgccacc	gtcgtcggtt	ctcagcttcc	tttcgggctt	tgtagcagc	cggatctcag	240
tggtgggtgg	gggtggtgct	gagtgcgccc	gcaagcttgt	cgacggagct	cgaattcgga	300
gagtcgaagg	aaagcccgaa	acaatcgctg	gcctagagtc	accaccacca	ccaccacgag	360
ctcagcgccg	cggtcgaaca	gctgcctcga	gcttaagcct	tccgatatca	gccatggcct	420
tgctgcgtc	gtcggtagcc	agatctgggc	tgctcatgtg	ctggcggtcg	aatttagcag	480
cagcgggttc	tttcatacca	aggctatagt	cggtaccgga	acagcagcag	cagccatggg	540
tctagaccgg	acaggtacac	gaccgcaagc	ttaaactcgt	gtcgccaaag	aaagtatggg	600
gaaccgcgtg	gcaccagacc	agaagaatga	tgatgatgat	ggtgcatatg	gccagaacca	660
gaaccggcca	ggtagcgctc	gaggaactgc	ttcaactgac	cttggcgctc	cgtgggtctg	720
tcttcttact	actactacta	ccacgtatac	cggtcttggt	cttggccggg	ccaatcgag	780
ctccttgaga	aagttgactg	ctttagacag	tgacccact	ttgggtgccg	ccacttcacc	840
gtttttgaac	agcagcagag	tcgggatacc	acggatgcca	tatttcggcg	cagtgcagg	900
gaaatctgtc	acgtgggtga	aaccaacggc	ggtgaagtgg	caaaaacttg	tgcgtcgtc	960
agccctatgg	tgctacgggt	ataaagccgc	gtcacgggtc	gttttgatcg	atgttcagtt	1020
ttgcaacggg	cagtttgccc	tgatatctgt	cagcgatttc	atccagaatc	ggggcgatca	1080
ttttcgacgg	accgcaccac	caaaactagc	tacaagtcaa	aacggttgcca	gtcaaacggg	1140
actataagca	gtcgctaaag	taggtcttag	ccccgctagt	aaaacgtgcc	tggtcggtgg	1200
tctgcccgag	aatcgacgag	gatcgccccg	tccgcttga	gtacatccgt	gtcaaaactg	1260
tcgtcagtc	ggtgaataat	tttatcgctc	atatgtatat	agacgggtct	ttagctgctc	1320
ctagcggggc	aggcgaaaot	catgtaggca	cagttttgac	agcagtcagt	ccacttatta	1380
aaatagcgag	tatacatata	ctccttctta	aagttaaaca	aaattatttc	tagaggggaa	1440
ttgttatccg	ctcacaaatc	ccctatagt	agtcgtatta	atttcgcggg	atcgagatcg	1500
gaggaagaat	ttcaatttgt	tttaataaag	atctccctt	aacaataggc	gagtggttaa	1560
gggatatac	tcagcataat	taaagcgccc	tagctctagc	atctcgatcc	tctacgccgg	1620
acgcatcggt	gcccggcatc	ccggcgccac	aggtgcgggt	gctggcgcc	atatcgccga	1680
catcaccgat	ggggaagatc	tagagctagg	agatgcggcc	tgctgtagc	cggccgtagt	1740
ggccgcgggt	tccacgccaa	cgaccgcgga	tatagcggct	gtagtggcta	ccccctctag	1800
gggtcgccca	cttcgggctc	ctgagcgctt	gtttcgccgt	gggtatgggt	gcaggccccg	1860
tgccgggggg	actgtggggc	gccatctcct	tgcatgcacc	cccgagcggg	gaagcccag	1920
tactcgcgaa	caaagccgca	cccataccac	cgtccggggc	accggccccc	tgacaaccgg	1980
cggtagagga	acgtacgtgg	attccttgcg	gcggcggtgc	tcaacggcct	caacctacta	2040
ctgggctgct	tcctaatagc	ggagtcgcat	aagggagagc	gtcgagatcc	cggacacccat	2100
taaggaacgc	cgccgccaag	agttgcggga	gttggtgat	gacccgacga	aggattacgt	2160
cctcagcgta	ttccctctcg	cagctctagg	gcctgtggta	cgaatggcgc	aaaacctttc	2220
gcggtagggc	atgatagcgc	ccggaagaga	gtcaattcag	ggtggtgaat	gtgaaaccag	2280
taacgttata	cgatgtcgca	gcttaccgcg	ttttggaaag	cgccataccg	tactatcgcg	2340
ggccttctct	cagttaagtc	ccaccactta	cactttgggt	attgcaatat	gctacagcgt	2400
gagtatgccg	gtgtctctta	tcagaccggt	tcccgcgtgg	tgaaccaggc	cagccacggt	2460
tctgcgaaaa	cgccgggaaa	agtggaaagc	gcgatggcgg	ctcataccgg	cacagagaat	2520
agtcctggca	aggcgccacc	acttggtccg	gtcgggtgca	agacgctttt	gcgccttttt	2580
tcaccttcgc	cgctaccggc	agctgaatta	cattcccac	cgctgggcac	aacaactggc	2640
gggcaaacag	tcgttgctga	ttggcggtgc	cacctccagt	ctggccctgc	acgcgcgctc	2700
tcgacttaat	gtaaggggtg	gogcaccgtg	ttgttgaccg	cccgtttgtc	agcaacgact	2760
aaccgcaacg	gtggagggtc	gaccgggacg	tgccggcgag	gcaaattgtc	gcggcgatta	2820
aatctcgcg	cgatcaactg	ggtgccagcg	tggtgggtgc	gatggtagaa	cgaagcgggc	2880
tcaagcctg	taaagcggcg	cgtttaacag	cgccgcta	ttagagcgcg	gctagttagc	2940
ccacgctcgc	accaccacag	ctaccatctt	gcttcgccgc	agcttcggac	atttcgccgc	3000
gtgcacaatc	ttctcgcgca	acgcgtcagt	gggtgatca	ttactatcc	gctggatgac	3060
caggatgcc	ttgctgtgga	agctgcctgc	actaatgttc	cacgtgttag	aagagcgcg	3120
tgcgagtc	cccgaactag	aattgatagg	cgacctactg	gtcctacggt	aacgacacct	3180
tcgacggag	tgattacaag	cgccgttatt	tcttgatgtc	tctgaccaga	cacccatcaa	3240
cagttattat	ttctcccatg	aagacgggac	gcgactgggc	gtggagcatc	tggtcgcat	3300
gccgcaataa	agaactacag	agactggctc	gtgggtagtt	gtcataataa	aagagggtac	3360
ttctgccatg	cgctgacccg	ccactcgtag	accagcgtaa	gggtcaccag	caaatcgcg	3420
tgtagcggg	cccattaagt	tctgtctcgg	cgctctgcg	tctggctggc	tggcataaat	3480
atctcactcg	caatcaaat	cccagtggtc	gttagcgcg	acaatcgccc	gggtaattca	3540
agacagagcc	gcgcagacgc	agaccgaccg	accgtattta	tagagtgagc	gttagtttaa	3600
cagccgatag	cggaaacggga	aggcgactgg	agtgcctatg	ccggttttca	acaaacctag	3660

caaagtctga	atgagggcat	cgttcccaact	gcgatgctgg	gtcgggtatc	gccttgccct	3720
tccgctgacc	tcacgggtaca	ggccaaaagt	tgtttggtac	gtttacgact	tactcccgta	3780
gcaaggggtga	cgctacgacc	ttgccaaacga	tcagatggcg	ctgggcgcaa	tgcgcgccat	3840
taccgagttcc	gggctgcgcg	ttggtgcgga	catctcggtg	gtgggatacg	acgataccga	3900
aacggttgct	agtctaccgc	gaccgcggtt	acgcgcggtg	atggctcagg	cccgcgcgc	3960
aaccacgcct	gtagagccat	caccctatgc	tgctatggct	agacagctca	tggtatatcc	4020
cgccgttaac	caccatcaaaa	caggattttc	gcctgctggg	gcaaaccagc	gtggaccgct	4080
tgctgcaact	ctctcagggc	tctgtcgagt	acaatatagg	gcggcaattg	gtggtagttt	4140
gtcctaaaag	cggacgaccc	cgtttggtcg	cacctggcga	acgacgttga	gagagtcccg	4200
caggcggtga	agggcaatca	gctgttgccc	gtctcactgg	tgaaaagaaa	aaccaccctg	4260
gcgcccata	cgcaaaccgc	ctctccccgc	gcgttgcccg	gtccgccact	tcccgtagt	4320
cgacaacggg	cagagtgcac	acttttcttt	ttggtgggac	cgcggttat	gcgtttggcg	4380
gagacggcg	attcattaat	cactctggca	cgacaggttt	cccactgga	cccactgga	4440
aagcgggag	tgagcgcaac	gcaattaatg	taagttagct	cactcattag	gcaccgggat	4500
taagtaatta	cgctcgaccgt	gctgtccaaa	gggctgacct	ttcgcccgct	actcgcggtg	4560
cgtttaattac	attcaatcga	gtgagtaatc	cgtaggccta	ctcgaccgat	gcccttgaga	4620
gccttcaacc	cagtcagctc	cttccgggtg	gcgcggggca	tgactatcgt	cgccgcactt	4680
atgactgtct	tctttatcat	gagctggcta	cggaactct	cggaagtgg	gtcagtcgag	4740
gaaggccacc	cgcgccccgt	actgatagca	gcggcgtaga	tactgacaga	agaaatagta	4800
gcaactcgta	ggcaggtgct	cggcagcgct	ttcggtcatt	ttcggcgagg	accgctttcg	4860
ctggagcgcg	acgatgatcg	gcctgtcgct	tgcggtattc	cggtgagcat	cctgtccacg	4920
gccgtcgcg	gaccagtaa	aagccgctcc	tgccgaaagc	gacctcgcg	tgctactagc	4980
cggacagcga	acgccataag	ggaatcttgc	acgccctcgc	tcaagccttc	gtcactgggtc	5040
ccgccaccaa	acgtttcggc	gagaagcagg	ccattatcgc	cgccatggcg	gccccacggg	5100
ccttagaacg	tgccggagcg	agttcggaag	cagtgaccag	ggcggtggtt	tgcaaagccg	5160
ctcttcgtcc	ggtaatatcg	ccggtaccgc	cggggtgccc	tgcgcatgat	cggtgctcgt	5220
tcgttgagga	cccggctagg	ctggcggggt	tgccctactg	gttagcagaa	tgaatcaccg	5280
atacgcgagc	gaacgtgaag	acgcgtacta	gcacgaggac	agcaactcct	gggcccgatcc	5340
gaccgcccc	acgggaatgac	caatcgtctt	acttagtggt	tatgcgctcg	cttgcaacttc	5400
cgactgctgc	tgcaaaacgt	ctgcgacctg	agcaacaaca	tgaatggtct	tccggtttccg	5460
tgtttcgtaa	agtcctggaa	cgcggaagtc	agcgccctgc	gctgacgacg	acgttttgca	5520
gacgctggac	tcgttggtgt	acttaccaga	agccaaaggc	acaaagcatt	tcagaccttt	5580
gcgccttcag	tcgcgggacg	accattatgt	tccggtctgt	catcgagga	tgctgctggc	5640
taccctgtgg	aacacctaca	tctgtattaa	cgaagcgctg	gcattgaccc	tgagtgattt	5700
tggtaatata	aggcctagac	gtagcgtcct	acgacgaccg	atgggacacc	ttgtggatgt	5760
agacataatt	gcttcgcgac	cgtaactggg	actcactaaa	ttctctgggt	ccgcgcgcatc	5820
cataccgcca	gttggtttacc	ctcacaacgt	tccagtaacc	gggcatgttc	atcatcagta	5880
accgctatcg	tgagcatcct	aagagacatc	ggcgcgtag	gtatggcggt	caacaaatgg	5940
gagtggtgca	aggctattgg	cccgtacaag	tagtagtcat	tgggcatagc	actcgtagga	6000
ctctcggttc	atcggtatca	ttacccccat	gaacagaaat	cccccttaca	cggaggcatc	6060
agtgacaaa	caggaaaaaa	ccgcccttaa	catggcccg	gagagcaaag	tagccatagt	6120
aatgggggta	cttgtcttta	gggggaatgt	gcctccgtag	tcactgggtt	gtcctttttt	6180
ggcggggaatt	gtaccgggag	tttatcagaa	gccagacatt	aacgcttctg	gagaaactca	6240
acgagctgga	cgcggatgaa	caggcagaca	tctgtgaatc	gcttcacgac	cacgctgatg	6300
aaatagctct	cggctctgtaa	ttgcgaagac	ctctttgagt	tgctcgacct	gcgcctactt	6360
gtccgtctgt	agacacttag	cgaagtgtcg	gtgcgactac	agctttaccg	cagctgcctc	6420
gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	agctcccgga	gacggtcaca	6480
gcttgctctg	aagcggatgc	tcgaaatggc	gtcgacggag	cgcgcaaagc	cactactgcc	6540
acttttgagg	actgtgtacg	tcgagggcct	ctgccagtgt	cgaacagaca	ttcgccctacg	6600
cgggagcaga	caagcccgtc	aggcgcgctg	agcggtgtgt	ggcggtgtgc	ggggcgagc	6660
catgaccag	tcacgttagcg	atagcggagt	gtatactggc	gccctcgctc	gttcgggagc	6720
tcccgcgag	tcgcccacaa	ccgcccacag	cccccgctcg	gtactgggtc	agtgcacgc	6780
tatcgctca	catatgaccg	ttactatgc	ggcatcagag	cagattgtac	tgagagtga	6840
ccatatatgc	gggtgtgaaat	accgcacaga	tgcgtaagga	gaaaataaccg	catcaggcgc	6900
aattgatacg	ccgtagtctc	gtctaacaatg	actctcaogt	ggtatatacg	ccacacttta	6960
tggcgtgtct	acgcattcct	cttttatggc	gtagtcccg	tcttcgctt	cctcgctcac	7020
tgactcgctg	cgcctcgctg	ttcggtgctg	gcgagcggtg	tcagctcact	caaaggcggt	7080
aatacggtta	tccacagaat	agaaggcgaa	ggagcgagtg	actgagcgac	gcgagccagc	7140
aagccgacgc	cgctcgccat	agtcgagtga	gtttccgcca	ttatgccaat	aggtgtctta	7200
caggggataa	cgcaggaaaag	aacatgtgag	caaaaggcca	gcaaaaggcc	aggaaccgta	7260
aaaaggccgc	gttgctggcg	ttttccata	ggctccgccc	gtccctatt	gcgtcctttc	7320

ttgtacactc	gttttccggt	cgttttccgg	tccttggcat	ttttccggcg	caacgaccgc	7380
aaaaaggtat	ccgaggcg	ccctgacgag	catcacaaaa	atcgacgctc	aagtcagagg	7440
tggcgaaacc	cgacaggact	ataaagatac	caggcggttc	cccctggaag	ctccctcgtg	7500
gggactgctc	gtagtgtttt	tagctgcgag	ttcagttctc	accgcttttg	gctgtcctga	7560
tatttctatg	gtccgcaaag	ggggaccttc	gagggagcac	cgctctcctg	ttccgaccct	7620
gccgcttacc	ggatacctgt	ccgcctttct	cccttcggga	agcgtggcgc	tttctcatag	7680
ctcacgctgt	aggtatctca	gcgagaggac	aaggctggga	cggcgaaatg	cctatggaca	7740
ggcggaaga	gggaagccct	tcgcaccgcg	aaagagtatc	gagtgcgaca	tccatagagt	7800
gttcggtgta	ggtcggttcg	tccaagctgg	gctgtgtgca	cgaaccccc	gttcagcccc	7860
accgctgcgc	cttatccggt	aactatcgtc	ttgagtccaa	caagccacat	ccagcaagcg	7920
aggttcgacc	cgacacacgt	gcttgggggg	caagtccggc	tggcgacgcg	gaataggcca	7980
ttgatagcag	aactcagggt	cccggtaaga	cacgacttat	cgccactggc	agcagccact	8040
ggtaacagga	ttagcagagc	gaggtatgta	ggcgtgtcta	cagagtctct	gaagtgggtg	8100
gggccattct	gtgctgaata	gcggtgaccg	tcgtcgggtg	ccattgtcct	aatcgtctcg	8160
ctccatacat	ccgccacgat	gtctcaagaa	cttcaccacc	cctaactacg	gctacactag	8220
aaggacagta	tttggtatct	gcgctctgct	gaagccagtt	accttcggaa	aaagagttgg	8280
tagctcttga	tccggcaaac	ggattgatgc	cgatgtgatc	ttcctgtcat	aaacataga	8340
cgcgagacga	cttcggtcaa	tggaaagcct	tttctcaacc	atcgagaact	aggccggttg	8400
aaaccaccgc	tggtagcggg	ggtttttttg	tttgcaagca	gcagattacg	cgcagaaaaa	8460
aaggatctca	agaagatcct	ttgatctttt	ctacgggggc	tttggtggcg	accatcgcca	8520
ccaaaaaac	aaacgttcgt	cgtctaattg	gcgtcttttt	ttcctagagt	tcttctagga	8580
aactagaaaa	gatgccccag	tgacgctcag	tggaaacgaa	actcacgtta	agggattttg	8640
gtcatgagat	tatcaaaaag	gatcttcacc	tagatccttt	taaattaaaa	atgaagtttt	8700
actgcgagtc	accttgcttt	tgagtgcaat	tccctaaaa	cagtactcta	atagtttttc	8760
ctagaagtgg	atctagaaaa	atttaatttt	tacttcaaaa	aaatcaatct	aaagtatatata	8820
tgagtaaaact	tggcttgaca	gttaccaatg	cttaatcagt	gaggcaccta	ctcagcgat	8880
ctgtctattt	cgttcatcca	tttagttaga	tttcatatat	actcatttga	accagactgt	8940
caatggttac	gaattagtca	ctccgtggat	agagtcgcta	gacagataaa	gcaagtaggt	9000
tagttgcctg	actccccgct	gtgtagataa	ctacgatacg	ggagggccta	ccatctggcc	9060
ccagtgtctg	aatgataccg	cgagacccac	gctcaccggc	atcaacggac	tgaggggag	9120
cacatctatt	gatgtctatg	ctcccgaa	ggtagaccgg	ggtcacgacg	ttactatggc	9180
gctctgggtg	cgagtggcgc	tccagattta	tcagcaataa	accagccagc	cgggaaggcc	9240
gagcgagaa	gtggctcctg	aactttatcc	gcctccatcc	agtctattaa	ttggtgccc	9300
aggtctaaat	agtcgttatt	tggctcggtc	gccttcccgg	ctcgcgtctt	caccaggacg	9360
ttgaaatagg	cggaggtagg	tcagataatt	aacaacggcc	gaagctagag	taagttagttc	9420
gccagttaat	agtttgcgca	acgttggtgc	cattgtgcga	ggcatcgtgg	tgtcacgctc	9480
gtcgtttggg	atggcttcat	cttcgatctc	attcatcaag	cgggtcaatta	tcaaacgcgt	9540
tgcaacaacg	gtaacgcagt	ccgtagacc	acagtgcgag	cagcaaacca	taccgaagta	9600
tcagctccgg	ttcccaacga	tcaaggcgag	ttacatgatc	ccccatgttg	tgcaaaaaag	9660
cggtagctc	cttcggtcct	ccgatcgttg	tcagaagtaa	agtcgaggcc	aagggttgct	9720
agttccgctc	aatgtactag	ggggtacaac	acgttttttc	gccaatcgag	gaagccagga	9780
ggctagcaac	agtcttcatt	gttgcccgca	gtgttatcac	tcatggttat	ggcagcactg	9840
cataattctc	ttactgtcat	gccatccgta	agatgctttt	ctgtgactgg	tgagtactca	9900
caaccggcgt	cacaatagtg	agtaccaata	ccgtcgtgac	gtattaagag	aatgacagta	9960
cggtaggcat	tctacgaaaa	gacactgacc	actcatgagt	accaagtcac	tctgagaata	10020
gtgtatgcgg	cgaccgagtt	gctcttgccc	ggcgtcaata	cgggataata	ccgcgccaca	10080
tagcagaact	ttaaaagtgc	tgggttcagta	agactcttat	cacatacgcc	gctggctcaa	10140
cgagaacggg	ccgcagttat	gccctattat	ggcgcggtgt	atcgtcttga	aattttcacg	10200
tcatcattgg	aaaacgttct	tccggggcgaa	aactctcaag	gatcttaccg	ctggtgagat	10260
ccagttcgat	gtaacccact	cgtgcaccca	actgatcttc	agtagtaacc	ttttgcaaga	10320
agccccgctt	ttgagagttc	ctagaatggc	gacaactcta	ggtcaagcta	cattgggtga	10380
gcacgtgggt	tgactagaag	agcatctttt	actttcacca	gcgtttctgg	gtgagcaaaa	10440
acaggaaggc	aaaatgccgc	aaaaaaggga	ataaggcgca	cacggaaatg	ttgaatactc	10500
tcgtagaaaa	tgaagtgggt	cgcaaaagacc	cactcgtttt	tgtccttcgg	ttttacggcg	10560
ttttttccct	tattcccgtc	gtgcctttac	aacttatgag	atactcttcc	tttttcaata	10620
ttattgaagc	atttatcagg	gttattgtct	catgagcgga	tacatatttg	aatgtattta	10680
gaaaaataaa	caaatagggg	tatgagaagg	aaaaagttat	aataaacttc	taaatagtcc	10740
caataacaga	gtactcgcct	atgtataaac	ttacataaat	ctttttattt	gtttatcccc	10800
ttccgcgcac	atttccccga	aaagtgccac	ctgaaattgt	aaacgttaat	attttgttta	10860
aattccgctt	aaatttttgt	taaatcagct	cattttttta	aaggcgctg	taaaggggct	10920
tttcacgggt	gactttaaca	tttgcaatta	taaaacaatt	ttaagcgcaa	tttaaaaaa	10980

```

atttagtcga gtaaaaaatt ccaataggcc gaaatcggca aaatccctta taaatcaaaa 11040
gaatagaccg agatagggtt gagggttgtt ccagtttggg acaagagtcc actattaaag 11100
ggttatccgg ctttagccgt tttagggaat atttagtttt cttatctggc tctatcccaa 11160
ctcacaaaaa ggtcaaacct tgttctcagg tgataatttc aacgtggact ccaacgtcaa 11220
agggcgaaaa accgtctatc agggcgatgg cccactacgt gaaccatcac cctaatacaag 11280
ttttttgggg tcgaggtgcc ttgcacctga ggttgcagtt tcccgttttt tggcagatag 11340
tcccgtacc gggatgatgca cttggtagtg ggattagttc aaaaaacccc agctccacgg 11400
gtaaagcact aaatcggaa cctaaaggga gcccccgatt tagagcttga cggggaaagc 11460
cggcgaaacgt ggcgagaaag gaagggaaga aagcgaaagg catttcgtga tttagccttg 11520
ggatttccct cgggggctaa atctcgaact gccctttcg gccgcttgca ccgctctttc 11580
cttccctttc ttcgcttttc agcgggcgct agggcgctgg caagtgtagc ggtcacgctg 11640
cgcgtaacca ccacaccgc cgcgcttaac gcgcgcgtac agggcgcgct ccattcgcca 11700
tcgcccgcga tcccgcgacc gttcacatcg ccagtcgcgac gcgcattggg ggtgtgggag 11760
gcgcgaatta cgcggcgatg tcccgcgcag ggtaagcggg 11800

```

<210> 33

<211> 2504

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: plasmide
d'expression pT7-7

<400> 33

```

aattctcatg tttgacagct tatcatcgat gataagcttg ggctgcaggt cgactctaga 60
ggatccccgg gcgcgaattc tagccatag tatatctcct tcttaaagtt aaacaaaatt 120
atttctagag ggaaaccggt gtggtctccc tatagttagt cgtattaatt tcgaagtcta 180
tcagaagttc gaatcgctgg gctcgcgcg tttcggtgat gacggtgaaa acctctgaca 240
catgcagctc ccggagacgg tcacagcttg tctgtaagcg gatgcggga gcagacaagc 300
ccgtcagggc gcgtcagcgg gtgtcggggc gcagccatga cccagtcacg 360
tagcgatagc ggagtgtata tactggctta actatgcggc atcagagcag attgtactga 420
gagtgcacca taggaagatc ttccggaaga tcttcctatg cgggtgtgaaa taccgcacag 480
atgcgtaagg agaaaatacc gcatcaggcg ctcttccgct tctcgcgtca ctgactcgt 540
gcgctcggtc gttcggctgc ggcgagcgg atcagctcac tcaaaggcgg taatacgggt 600
atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc 660
caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggtcccgcc cccctgacga 720
gcatcacaaa aatcgacgct caagtacagag gtggcgaaac ccgacaggac tataaagata 780
ccaggcgttt cccctggaa gctccctcgt gcgctctcct gttccgaccc tgcgcttac 840
cggatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcaat gctcacgctg 900
taggtatctc agttcggtgt aggtcggtcg ctccaagctg ggctgtgtgc acgaaccccc 960
cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca acccggttaag 1020
acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag cgaggatatg 1080
aggcgggtgt acagagttct tgaagtgggt gcctaactac ggctacacta gaaggacagt 1140
atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg gtagctcttg 1200
atccggcaaa caaaccaccg ctggtagcgg tgggtttttt gtttgcaagc agcagattac 1260
gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggg ctgacgctca 1320
gtggaacgaa aactcacggt aagggtatgt ggtcatgaga ttatcaaaaa ggatcttcac 1380
ctagatcctt ttaattcttg aagacgaaag ggcctcgtga tacgcctatt tttataggtt 1440
aatgtcatga taataatggt ttcttagacg tcaggtggca cttttcgggg aaatgtgctg 1500
ggaaccccta tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa 1560
taaccctgat aaatgcttca ataattattg aaaaggaaga gtatgagtat tcaacatttc 1620
cgtgtcgccc ttattccctt ttttgccgca ttttgccctc ctgttttttg tcacccagaa 1680
acgctggtga aagtaaaaga tgctgaagat cagttgggtg cacgagtggg ttacatcgaa 1740
ctggtactca acagcggtaa gatccttgag agttttcgcc ccgaagaacg ttttccaatg 1800
atgagcactt ttaaagttct gctatgtggc gcggtattat cccgtgttga cgcggggcaa 1860
gagcaactcg gtcgccgcat acactattct cagaatgact tgggtgagta ctaccagtc 1920
acagaaaagc atcttacgga tggcatgaca gtaagagaat tatgcagtgc tgccataacc 1980
atgagtgata acactgcggc caacttactt ctgacaacga tcggaggacc gaaggagcta 2040
accgcttttt tgcacaacat gggggatcat gtaactcgcc ttgatcgttg ggaaccggag 2100

```



```

ctgaatgaag ccataccaaa cgacgagcgt gacaccacga tgccctgtagc aatggcaaca 2160
acgttgcgca aactattaac tggcgaaacta cttactctag cttcccgga acaattaata 2220
gactgggatg aggcggataa agttgcagga ccacttctgc gtcggccct tccggctggc 2280
tggtttattg ctgataaatc tggagccggt gacgctgggt ctgcgggtat cattgcagca 2340
ctggggccag atggtaagcc ctcccgatc gtagttatct acacgacggg gagtcaggca 2400
actatggatg aacgaaatag acagatcgct gagataggtg cctcactgat taagcattgg 2460
taactgtcag accaagttta ctcatatata ctttagattg attt 2504

```

<210> 34

<211> 813

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: système
d'expression codant pour la protéine de fusion
GST-DP-TME1

<400> 34

```

atgtccccta tactaggtta ttggaaaatt aagggccttg tgcaaccac tcgacttctt 60
ttggaatata ttgaagaaaa atatgaagag catttgtag agcgcgatga aggtgataaa 120
tggcgaaaca aaaagtttga attgggtttg gagtttccca atcttcctta ttatattgat 180
ggtgatgta aattaacaca gtctatggcc atcatagctt atatactga caagcacaac 240
atgttgggtg gttgtccaaa agagcgtgca gagatttcaa tgcttgaagg agcgggtttg 300
gatattagat acggtgtttc gagaattgca tatagtaaag actttgaaac tctcaaagtt 360
gattttctta gcaagctacc tgaaatgctg aaaatgttcg aagatcgttt atgtcataaa 420
acatatata atggtgatca tgtaacccat cctgacttca tggtgtatga cgctcttgat 480
gttgttttat acatggaccc aatgtgcctg gatgcgttcc caaaattagt ttgttttaaa 540
aaacgtattg aagctatccc acaaattgat aagtacttga aatccagcaa gtatatagca 600
tggcctttgc agggctggca agccacggtt ggtgggtggc accatcctcc aaaatcggat 660
ctgtctggtg gtggtggtg tctggttccg cgtggatcgc acccgatcgc tgggtgctcag 720
tggggtgttc tgggtggtat cgcttacttc tctatggtg gtaactgggc taaagttctg 780
gttgttctgc tgctgttcgc tgggtgttgac gct 813

```

<210> 35

<211> 513

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: système
d'expression codant pour la protéine de fusion
TrX-DP-TME1

<400> 35

```

atgagcgata aaattattca cctgactgac gacagttttg acacggatgt actcaaagcg 60
gacggggcga tcctcgctga tttctgggca gagtgggtgc gtccgtgcaa aatgatcgcc 120
ccgattcttg atgaaatcgc tgacgaatat cagggcaaac tgaccgttgc aaaactgaac 180
atcgatcaaa accctggcac tgcgcccgaat tatggcatcc gtggatatcc gactctgctg 240
ctgttcaaaa acggtgaagt ggcggcaacc aaagtgggtg cactgtctaa aggtcagttg 300
aaagagtccc tcgacgctaa cctggccggt tctggttctg gatctccaaa atcggtatctg 360
tctggtggtg gtggtggtct ggttccgcgt ggatccgacc cgatcgctgg tgctcactgg 420
ggtgttcttg ctggtatcgc ttacttctct atggttggtg actgggctaa agttctggtt 480
gttctgctgc tgttcgctgg tgttgacgct tag 513

```

<210> 36

<211> 117

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: système
d'expression codant pour la protéine de fusion
M-DP-TME1

<400> 36

atggaccgga tcgctgggtgc tcaactggggt gttctgggtg gtatcgctta cttctctatg 60
atggtaact gggctaaagt tctggttgtt ctgctgctgt tcgctgggtg tgacgct 117

<210> 37

<211> 795

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: système
d'expression codant pour la protéine de fusion
GST-DP-TME2

<400> 37

atgtccccta tactaggtta ttggaaaatt aagggccttg tgcaaccac tcgacttctt 60
ttggaatatac ttgaagaaaa atatgaagag catttgtatg agcgcgatga aggtgataaa 120
tggcgaaaca aaaagtttga attgggtttg gagtttccca atcttcctta ttatattgat 180
ggtgatgta aattaacaca gtctatggcc atcatacgtt atatacgtga caagcacaac 240
atgttgggtg gttgtccaaa agagcgtgca gagatttcaa tgcttgaagg agcggttttg 300
gatattagat acggtgtttc gagaattgca tatagtaaag actttgaaac tctcaaagtt 360
gattttctta gcaagctacc tgaaatgctg aaaatgttcg aagatcgttt atgtcataaa 420
acatatattaa atggtgatca tgtaacccat cctgacttca tgttgatga cgctcttgat 480
gttgttttat acatggaccc aatgtgcctg gatgcgttcc caaaattagt ttgttttaaa 540
aaacgtattg aagctatccc acaaattgat aagtacttga aatccagcaa gtatatagca 600
tggcctttgc agggctggca agccacgttt ggtgggtggc accatcctcc aaaatcggat 660
ctgtctgggtg gtgggtgggtg tctggttccg cgtggatccg acccggaata cgttgttctg 720
ctgttctctg tcgtggctga cgtcgtgtt tgctcttgcc tgtggatgat gctgctgac 780
tctcaggctg aagct 795

<210> 38

<211> 486

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: système
d'expression codant pour la protéine de fusion
TrX-DP-TME2

<400> 38

atgagcgata aaattattca cctgactgac gacagttttg acacggatgt actcaaagcg 60
gacggggcga tcctcgtcga tttctgggca gagtgggtgc gtccgtgcaa aatgatcgcc 120
ccgattctgg atgaaatcgc tgacgaatat cagggcaaac tgaccgttgc aaaactgaac 180
atcgatcaaaa accctggcac tgcgcgcaaa tatggcatcc gtggtatccc gactctgctg 240
ctgttcaaaa acggtgaagt ggcggcaacc aaagtgggtg cactgtctaa aggtcagttg 300
aaagagttcc tcgacgctaa cctggccggg tctggttctg gatctgatct gtctgggtgg 360
ggtgggtggc tggttccgcg tggatccgac ccggaatacg ttgttctgct gttcctgctg 420
ctggctgacg ctctgttttg ctcttgctg tggatgatgc tgctgatctc tcaggctgaa 480
gcttag 486

<210> 39
 <211> 99
 <212> ADN
 <213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: système
 d'expression codant pour la protéine de fusion
 M-DP-TME2

<400> 39

atgggacccgg aatacgttgt tctgctgttc ctgctgctgg ctgacgctcg tgtttgctct 60
 tgccctgtgga tgatgctgct gatctctcag gctgaagct 99

<210> 40

<211> 5082

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: vecteur
 d'expression pGEXKT-dp-Pt (TME1)

<400> 40

acgttatcga ctgcacgggtg caccaatgct tctggcgtca ggcagccatc ggaagctgtg 60
 gtatggctgt gcaggctcgt aatcactgca taattcgtgt cgctcaaggc gcactcccgt 120
 tctggataat gttttttgcg ccgacatcat aacggttctg gcaaatattc tgaaatgagc 180
 tgttgacaat taatcatcgg ctcgtataat gtgtggaatt gtgagcggat aacaatttca 240
 cacaggaaac agtattcatg tcccctatac taggttattg gaaaattaag ggccttgtgc 300
 aacccactcg acttcttttg gaatatcttg aagaaaaata tgaagagcat ttgtatgagc 360
 gcgatgaagg tgataaatgg cgaacacaaa agtttgaatt gggtttggag tttcccaatc 420
 ttccttatta tattgatggg gatgttaaata taacacagtc tatggccatc atacgttata 480
 tagctgacaa gcacaacatg ttgggtgggt gtccaaaaga gcgtgcagag atttcaatgc 540
 ttgaaggagc ggttttggat attagatacg gtgtttcgag aattgcatat agtaaagact 600
 ttgaaactct caaagttgat tttcttagca agctacctga aatgctgaaa atgttcgaag 660
 atcgttttat tcataaaaca tatttaaatg gtgatcatgt aacccatcct gacttcgatg 720
 tgatgacgc tcttgatggt gttttataca tggacccaat gtgcctggat gcgttcccaa 780
 aattagtttg ttttaaaaaa cgtattgaag ctatcccaca aattgataag tacttgaaat 840
 ccagcaagta tatagcatgg cctttgcagg gctggcaagg cacgtttggg ggtggcgacc 900
 atcctccaaa atcggatctg tctggtgggt gtggtgggtc ggttccgcgt ggatccgacc 960
 cgatcgctgg tgctcactgg ggtgttctgg ctggtatcgc ttacttctct atggttggtg 1020
 actgggctaa agttctgggt gttctgctgc gtctcgctgg tgttgacgct taggaattca 1080
 tcgtgactga ctgacgatct gcctcgcgcg tttcggtgat gacggtgaaa acctctgaca 1140
 catgcagctc ccggagacgg tcacagcttg tctgtaagcg gatgccggga gcagacaagc 1200
 ccgtcagggc gcgtcagcgg gtgttgggcg gtgtcggggc gcagccatga cccagtcacg 1260
 tagcgatagc ggagtgatata attcttgaag acgaaagggc ctctgatac gcctattttt 1320
 ataggttaat gtcatagata taatgggttc ttagacgtca ggtggcactt ttcggggaaa 1380
 tgtgcgcgga acccctattt gtttattttt taaaatacat tcaaatatgt atccgctcat 1440
 gagacaataa ccttgataaa tgcttcaata atattgaaaa aggaagagta tgagtattca 1500
 acatttccgt gtcgccctta ttcctttttt tgcggcattt tgccttctg tttttgctca 1560
 cccagaaacg ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac gagtgggtta 1620
 catcgaaact gatctcaaca gcggtaagat ccttgagagt tttcgccccg aagaacgttt 1680
 tccaatgatg agcactttta aagttctgct atgtggcgcg gtattatccc gtgttgacgc 1740
 cgggcaagag caactcgggt gccgcataca ctattctcag aatgacttgg ttgagtactc 1800
 accagtacac gaaaagcatc ttacggatgg catgacagta agagaattat gcagtgtcgc 1860
 cataaccatg agtgataaca ctgcgcccaa cttacttctg acaacgatcg gaggaccgaa 1920
 ggagctaacc gcttttttgc acaacatggg ggatcatgta actcgccttg atcgttggga 1980
 accggagctg aatgaagcca taccaaacga cgagcgtgac accacgatgc ctgcagcaat 2040
 ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt cccggcaaca 2100
 attaatagac tggatggagg cggataaagt tgcaggacca cttctgctgc cggcccttcc 2160

```

ggctggctgg tttattgctg ataaatctgg agccgggtgag cgtgggtctc gcggtatcat 2220
tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca cgacggggag 2280
tcaggcaact atggatgaac gaaatagaca gatcgctgag ataggtgcct cactgattaa 2340
gcattggtaa ctgtcagacc aagtttactc atatatactt tagattgatt taaaacttca 2400
tttttaattt aaaaggatct aggtgaagat cctttttgat aatctcatga ccaaaatccc 2460
ttaacgtgag ttttcgttcc actgagcgtc agaccccgta gaaaagatca aaggatcttc 2520
ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac caccgctacc 2580
agcgggtggt tgtttgccgg atcaagagct accaactctt tttccgaagg taactggctt 2640
cagcagagcg cagataccaa atactgtcct tctagtgtag ccgtagttag gccaccactt 2700
caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac cagtggctgc 2760
tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt taccggataa 2820
ggcgccagcg tcggggtgaa cggggggttc gtgcacacag cccagcttgg agcgaacgac 2880
ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcgccacgc tcccgaagg 2940
gagaaaggcg gacaggtatc cggtgaagcg cagggtcgga acaggagagc gcacgagggg 3000
gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc acctctgact 3060
tgagcgtcga tttttgtgat gctcgtcagg gggcgggagc ctatggaaaa acgccagcaa 3120
cgcgcccttt ttacggttcc tggccttttg ctggcctttt gctcagatgt tctttcctgc 3180
gttatccctg gattctgtgg ataaccgtat taccgccttt gagttagctg ataccgctcg 3240
ccgcagccga acgaccgagc gcagcagctc agtgagcggag gaagcggaag agcgccctgat 3300
gcggatattt ctccctacgc atctgtgcgg tatttcacac cgcataaatt ccgacaccat 3360
cgaatggtgc aaaacctttc gcggtatggc atgatagcgc ccggaagaga gtcaattcag 3420
ggtggtgaat gtgaaaccag taacgttata cgatgtcgca gagtatgccg gtgtctctta 3480
tcagaccgtt tcccgcgtgg tgaaccaggc cagccacggt tctgcgaaaa cgcgggaaaa 3540
agtggaaagc gcgtggcggt agctgaatta cattcccaac cgcgtggcac acaactggc 3600
gggcaaacag tcgttgctga ttggcggtgc cacctccagt ctggccctgc acgcgccgtc 3660
gcaaatgtc gcggcgatta aatctcgcgc cgatcaactg ggtgccagcg tgggtggtgc 3720
gatggtagaa cgaagcgggc tcgaagcctg taaagcggcg gtgcacaatc ttctcgcgca 3780
acgcgtcagt gggctgatca ttaactatcc gctggatgac caggatgcca ttgctgtgga 3840
agctgcctgc actaatgttc cggcgttatt tcttgatgtc tctgaccaga caccatcaa 3900
cagtattatt ttctcccatg aagacggtac gcgactgggc gtggagcatc tggtcgcatt 3960
gggtcaccag caaatcgccg cgttagcggg ccattaaagt tctgtctcgg cgcgtctcg 4020
tctggctggc tggcataaat atctcactcg caatcaaatt cagccgatag cggaacggga 4080
aggcgactgg agtgccatgt ccggttttca acaaaccatg caaatgctga atgagggcat 4140
cgttccact gcgatgctgg ttgccaacga tcagatggcg ctggcgcaa tgccgcgccat 4200
taccgagtcc gggctgcgcg ttggtgcgga tatctcggtg gtgggatacg acgataccga 4260
agacagctca tgttatatcc cgccgttaac caccatcaaa caggattttc gcctgctggg 4320
gcaaaccagc gtggaccgct gtctgcaact ctctcagggc caggcggtga agggcaatca 4380
gctgttgccc gtctcactgg tgaaaagaaa aaccaccctg gcgccaata cgcaaaccgc 4440
ctctccccgc gcgttgcccg attcattaat gcagctggca cgacaggttt cccgactgga 4500
aagcgggagc tgagcgcaac gcaattaatg tgagttagct cactcattag gcacccagg 4560
ctttacactt tatgttcccg gctcgtatgt tgtgtggaat tgtgagcggg taacaatttc 4620
acacaggaaa cagctatgac catgattacg gattcactgg ccgtcgtttt acaacgtcgt 4680
gactgggaaa accctggcgt taccgaactt aatcgccctg cagcacatcc ccctttcgcc 4740
agctggcgta atagcgaaga ggcccgacc gatcgccctt cccaacagtt gcgcagcctg 4800
aatggcgaat ggcgctttgc ctggtttccg gcaccagaag cgggtgccgga aagctggctg 4860
gagtgcgac ttcttgaggc cgatactgtc gtcgtccctt caaactggca gatgcacggg 4920
tacgatgcgc ccactacac caacgtaacc tatccatta cggccaatcc gccgtttgtt 4980
cccacggaga atccgacggg ttgttactcg ctcacattta atgttgatga aagctggcta 5040
caggaaggcc agacgcgaat tatttttgat ggcgttgga tt 5082

```

<210> 41

<211> 5064

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: vecteur d'expression pGEXKT-dp-Pt (TME2)

<400> 41

acgttatcga	ctgcacgggt	caccaatgct	tctggcgctca	ggcagccatc	ggaagctgtg	60
gtatggctgt	gcaggctcgta	aatcactgca	taattcgtgt	cgctcaaggc	gcactcccgt	120
tctggataat	gttttttgcg	ccgacatcat	aacggttctg	gcaaataatc	tgaaatgagc	180
tgttgacaat	taatcatcgg	ctcgatataat	gtgtggaatt	gtgagcggat	aacaatttca	240
cacaggaaac	agtattcatg	tcccctatac	taggttattg	gaaaattaag	ggccttgtgc	300
aaccactcgc	acttcttttg	gaatatcttg	aagaaaaata	tgaagagcat	ttgtatgagc	360
gcatgaagg	tgataaatgg	cgaacaaaaa	agtttgaatt	gggtttggag	tttcccaatc	420
ttccttatta	tattgatggg	gatgttaaat	taacacagtc	tatggccatc	atacgttata	480
tagctgacaa	gcacaacatg	ttgggtgggt	gtccaaaaga	gcgtgcagag	atttcaatgc	540
ttgaaggagc	ggttttggat	attagatacg	gtgtttcgag	aattgcatat	agtaaagact	600
ttgaaactct	caaagttgat	tttcttagca	agctacctga	aatgctgaaa	atgttcgaag	660
atcgtttatg	tcataaaaaca	tatttaaatg	gtgcatgt	aacctatcct	gacttcatgt	720
tgtatgacgc	tcttgatgtt	gttttataca	tggacccaat	gtgcctggat	gcgttcccaa	780
aattagtttg	ttttaaaaaa	cgtattgaag	ctatcccaca	aattgataag	tacttgaaat	840
ccagcaagta	tatagcatgg	cctttgcagg	gctggcaagc	cacgtttggg	gggtggcgacc	900
atcctccaaa	atcggatctg	tctgggtggg	gtgggtgggt	gggtccgcgt	ggatccgacc	960
cgaataacgt	tgttctgctg	ttcctgctgc	tggctgacgc	tcgtgtttgc	tcttgccctgt	1020
ggatgatgct	gctgatctct	caggctgaag	cttaggaatt	catcgtgact	gactgacgat	1080
ctgcctcgcg	cgtttccggg	atgacgggtg	aaacctctga	cacatgcagc	tcccggagac	1140
ggtcacagct	tgtctgtaag	cggatgccgg	gagcagacaa	gcccgtcagg	gcgcgtcagc	1200
gggtgttggc	gggtgtcggg	gcgcagccat	gaccagtcga	cgtagcgata	gcggagtgtg	1260
taattcttga	agacgaaagg	gcctcgtgat	acgcctatct	ttatagggtg	atgtcatgat	1320
aataatgggt	tcttagacgt	cagggtggcag	ttttcgggga	aatgtgcgcg	gaacccttat	1380
ttgttttatt	ttctaataac	attcaaatat	gtatccgctc	atgagacaat	aacctgata	1440
aatgtttcaa	taataattga	aaaggaagag	tatgagtatt	caacatttcc	gtgtcgccct	1500
tattcccttt	tttgcggcat	tttgccttcc	tgtttttgct	caccagaaa	cgctgggtgaa	1560
agtaaaagat	gctgaagatc	agttgggtgc	acgagtgggt	tacatcgaa	tggatctcaa	1620
cagcggtaag	atccttgaga	gttttcgccc	cgaagaacgt	tttccaatga	tgagcacttt	1680
taaagttctg	ctatgtggcg	cggattatc	ccgtgttgac	gccgggcaag	agcaactcgg	1740
tcgccgcata	cactattctc	agaatgactt	ggttgagtac	tcaccagtca	cagaaaagca	1800
tcttaccggat	ggcatgacag	taagagaatt	atgcagtgtc	gccataacca	tgagtataaa	1860
cactgcggcc	aacttacttc	tgacaacgat	cggaggaccg	aaggagctaa	ccgctttttt	1920
gcacaacatg	ggggatcatg	taactcgcct	tgatcgttgg	gaaccggagc	tgaatgaagc	1980
cataccaaac	gacgagcgtg	acaccacgat	gcctgcagca	atggcaacaa	cgttgcgcaa	2040
actattaact	ggcgaactac	ttactctagc	ttcccggcaa	caattaatag	actggatgga	2100
ggcggataaa	gttgcaggac	cacttctgcg	ctcgccctt	ccggtgggt	ggtttattgc	2160
tgataaatct	ggagcgggtg	agcgtgggtc	tcgcggtatc	attgcagcac	tggggccaga	2220
tggtaagccc	tcccgatatc	tgattatcta	cacgacgggg	agtcaggcaa	ctatggatga	2280
acgaaataga	cagatcgctg	agataggtgc	ctcactgatt	aagcattggg	aactgtcaga	2340
ccaagtttac	tcatatatac	tttagattga	tttaaaactt	catttttaat	ttaaaaggat	2400
ctaggtgaag	atcctttttg	ataatctcat	gaccaaaatc	ccttaacgtg	agttttctgt	2460
ccactgagcg	tcagaccccc	tagaaaagat	caaaggatct	tcttgagatc	ctttttttct	2520
gcgcgtaatc	tgctgcttgc	aaacaaaaaa	accaccgcta	ccagcgggtg	tttgtttgcc	2580
ggatcaagag	ctaccaactc	tttttccgaa	ggtaactggc	ttcagcagag	cgcagatacc	2640
aaatactgtc	cttctagtgt	agccgtagtt	aggccaccac	ttcaagaact	ctgtagcacc	2700
gcctacatac	ctcgctctgc	taatcctgtt	accagtggct	gctgccagtg	gcgataagtc	2760
gtgtcttacc	gggttggact	caagacgata	gttaccggat	aaggcgcagc	ggtcgggctg	2820
aacggggggg	tcgtgcacac	agcccagctt	ggagcgaacg	acctacaccg	aactgagata	2880
cctacagcgt	gagctatgag	aaagcggcac	gcttcccga	gggagaaagg	cggacaggta	2940
tccggttaagc	ggcaggggtc	gaacaggaga	gcgcacgagg	gagcttccag	ggggaacgcg	3000
ctggatatct	tatagtcctg	tccgggttctg	ccacctctga	cttgagcgtc	gatttttctg	3060
atgctcgtea	ggggggcgga	gcctatggaa	aaacgccagc	aacgcggcct	ttttacggtt	3120
cctggccctt	tgctggcctt	ttgctcacat	gttctttcct	gcgttatccc	ctgattctgt	3180
ggataaccgt	attaccgcct	ttgagtgaag	tgataccgct	cgccgcagcc	gaacgaccga	3240
gcgcagcgag	tcagtgaagc	aggaagcgga	agagcgccct	atgcggtatt	ttctccttac	3300
gcactctgtc	ggtatttcac	accgcataaa	ttccgacacc	atcgaatggg	gcaaaacctt	3360
tcgcgggatg	gcagatagc	gcccggaaga	gagtaaatc	aggggtggtg	atgtgaaacc	3420
agtaacgtta	tacgatgtcg	cagagtatgc	cgggtgtctc	tatcagaccg	tttcccgcgt	3480
gggtgaaccag	gccagccacg	tttctgcgaa	aacgcgggaa	aaagtggaa	cggcgatggc	3540
ggagctgaat	tacattccca	accgcgtggc	acaacaactg	gcgggcaaac	agtcgttgct	3600
gattggcggt	gccacctcca	gtctggccct	gcacgcgccg	tcgcaaattg	tcgcggcgat	3660

```

taaattctgc gccgatcaac tgggtgccag cgtgggtggtg tcgatggtag aacgaagcgg 3720
cgtcgaagcc tgtaaagcgg cggtgcacaa tcttctcgcg caacgcgtca gtgggtgat 3780
cattaactat ccgctggatg accaggatgc cattgctgtg gaagctgcct gcactaatgt 3840
tccggcggtta tttcttgatg tctctgacca gacacccatc aacagtatta ttttctccca 3900
tgaagacggg acgcgactgg gcgtggagca tctgggtcgca ttgggtcacc agcaaatcgc 3960
gctgttagcg ggcccattaa gttctgtctc ggcgcgctctg cgtctggctg gctggcataa 4020
atatctcact cgcaatcaaa ttcagccgat agcggaaacgg gaaggcgact ggagtgccat 4080
gtccgggtttt caacaaacca tgcaaatgct gaatgagggc atcgttccca ctgcgatgct 4140
ggttgccaac gatcagatgg cgctgggctg aatgcgcgcc attaccgagt ccgggtgctg 4200
cgttgggtgcg gatattctcg tagtgggata cgacgatacc gaagacagct catgttatat 4260
cccgcggtta accaccatca aacaggattt tgcctgtctg gggcaaacca gcgtggaccg 4320
cttgctgcaa ctctctcagg gccaggcggt gaagggcaat cagctgttgc ccgtctcact 4380
ggtgaaaaga aaaaccaccc tggcgcccaa tacgcaaacc gcctctcccc gcgcgttggc 4440
cgattcatta atgcagctgg cagcagaggt tccccgactg gaaagcgggc agtgagcgca 4500
acgcaattaa tgtgagttag ctcaactcatt aggcacccca ggctttacac tttatgcttc 4560
cggctcgatg gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg 4620
accatgatta cggattoact ggccgtcggt ttacaacgtc gtgactggga aaaccctggc 4680
gttaccacaac ttaatcgctt tgcagcacat cccctcttcg ccagctggcg taatagcgaa 4740
gaggcccgca ccgatcgccc ttcccaacag ttgcgcagcc tgaatggcga atggcgcttt 4800
gcctgggtttc cggcaccaga agcgtgccc gaaagctggc tggagtgcga tcttcttgag 4860
gcccgaactg tcgtcgctcc ctcaaactgg cagtagcacg gttacgatgc gcccatctac 4920
accaacgtaa cctatcccat tacggtcaat ccgcggtttg tttccacgga gaatccgacg 4980
ggttgttact cgctcacatt taatgttgat gaaagctggc tacaggaagg ccagacgcga 5040
attatttttg atggcggttg aatt 5064

```

<210> 42

<211> 5918

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: vecteur
d'expression pET32a-dp-Pt (TME1)

<400> 42

```

atccggatat agttcctcct ttcagcaaaa aaccctctcaa gaccggttta gaggcccca 60
gggttatgc tagttattgc tcagcgggtg cagcagccaa ctacagcttc tttcgggctt 120
tgtagcagc cggatctcag cgggtgtggt ggtggtgctc gaggcgccgc gcaagcttgt 180
cgacggagct cgaattccta agcgtcaaca ccagcgaaca gcagcagaac aaccagaact 240
ttagccagc tacciaacct agagaagtaa gcgataccag ccagaacacc ccagttagca 300
ccagcgatcg ggtcggatcc acgcggaacc agaccaccac caccaccaga cagatccgat 360
tttgagatc cagaaccaga accggccagg ttagcgtcga ggaactcttt caactgacct 420
ttagacagtg caccactttt ggttgccgcc acttcaccgt ttttgaacag cagcagagtc 480
gggataccac ggatgccata tttcggcgca gtgccagggt tttgatcgat gttcagtttt 540
gcaacgggtc gtttgccctg atattcgtca gcgatttcat ccagaatcgg ggcgatcatt 600
ttgcacggac cgcaccactc tgcccagaaa tcgacgagga tcgccccgtc cgctttgagt 660
acatccgtgt caaaactgtc gtcagtcagg tgaataattt tatcgctoat atgtatatct 720
ccttcttaaa gttaaacaata attatttcta gaggggaatt gttatccgct cacaattccc 780
ctatagttag tcgtattaat ttcgcgggat cgagatcgat ctcgatcctc tacgccggac 840
gcacgtggc cggcatcacc ggcgccacag gtgcggttgc tggcgccat atcgccgaca 900
tcaccgatgg ggaagatcgg gtcgccact tcgggtcat gagcgcttg ttcggcggtg 960
gtatggtggc aggcctcggt gccgggggac tgttggggcg catctccttg catgcacct 1020
tccttgccgc ggcggtgctc aacggcctca acctactact gggctgcttc ctaatgcagg 1080
agtcgcataa gggagagcgt cgagatcccc gacaccatcg aatggcgcaa aacctttcgc 1140
ggtatggcat gatagcgccc ggaagagagt caattcaggg tggatgaatgt gaaaccagta 1200
acgttatagc atgtcgcaga gtagcgggt gtctcttacc agaccgtttc ccgctggtg 1260
aaccaggcca gccacgtttc tcgaaaaag cgggaaaaag tggaaagcgg gatggcggag 1320
ctgaattaca tttcccaacc cgtggcacaa caactggcgg gcaaacagtc gttgctgatt 1380
ggcggttgca cctccagtc ggcctgcac gcgcgctcgc aaattgtcgc ggcgattaaa 1440
tctcgcgccg atcaactggg tgccagcgtg gtggtgtcga tggtagaacg aagcggcgctc 1500

```

gaagcctgta	aagcggcggg	gcacaatcct	ctcgcgcaac	gcgtcagtg	gctgatcatt	1560
aactatccgc	tggatgacca	ggatgccatt	gctgtggaag	ctgcctgcac	taatgttccg	1620
gcgttatttc	ttgatgtctc	tgaccagaca	cccatcaaca	gtattatttt	ctcccatgaa	1680
gacgggtacgc	gactgggctg	ggagcatctg	gtcgcatctg	gtcaccagca	aatcgcgctg	1740
ttagcggggcc	cattaagttc	tgtctcggcg	cgtctgcgtc	tggctggctg	gcataaata	1800
ctcactcgca	atcaaattca	gccgatagcg	gaacgggaag	gcgactggag	tgccatgtcc	1860
ggttttcaac	aaaccatgca	aatgctgaat	gagggcatcg	ttccactgc	gatgctgggt	1920
gccaacgac	agatggcgct	gggcgcaatg	accggcatta	ccgagtcggg	gctgcgcgtt	1980
gggtgcggaca	tctcggtagt	gggatacgac	gataccgaag	acagctcatg	ttatatcccg	2040
ccgttaacca	ccatcaaaca	ggattttcgc	ctgctggggc	aaaccagcgt	ggaccgcttg	2100
ctgcaactct	ctcagggcca	ggcgggtgaag	ggcaatcagc	tggtgcccgt	ctcactgggt	2160
aaaagaaaaa	ccaccctggc	gcccaatacg	caaaccgcct	ctccccgcgc	gttgggccgat	2220
tcattaatgc	agctggcacg	acaggtttcc	cgactggaaa	gcgggcagtg	agcgcaacgc	2280
aattaatgta	agttagctca	ctcattagcg	accgggatct	cgaccgatgc	ccttgagagc	2340
cttcaacca	gtcagctcct	tccgggtggc	gcggggcatg	actatcgctc	ccgcacttat	2400
gactgtcttc	tttatcatgc	aactcgtagg	acaggtgccc	gcagcgctct	gggtcatttt	2460
cggcgaggac	cgctttcgtc	ggagcgcgac	gatgatcggc	ctgtcgcttg	cggtattcgg	2520
aatcttgac	gccctcgctc	aagccttcgt	cactgggtccc	gccaccaaac	gtttcggcga	2580
gaagcaggcc	attatcgccg	gcatggcggc	cccacgggtg	cgcatgatcg	tgctcctgtc	2640
gttgaggacc	ggcttaggct	ggcgggggtg	ccttactggg	tagcagaatg	aatcaccgat	2700
acgcgagcga	acgtgaagcg	actgctgctg	caaaaacgtc	gcgacctgag	caacaacatg	2760
aatgggtcttc	ggtttcctgt	tttcgtaaa	tctggaaaacg	cggaagtcag	cgccctgcac	2820
cattatgttc	eggatctgca	tcgcaggatg	ctgctggcta	ccctgtggaa	cacctacatc	2880
tgtattaacg	aagcgtggc	attgaccctg	agtgattttt	ctctgggtccc	gccgcatcca	2940
taccgccagt	tgtttaacct	cacaacgttc	cagtaaccgg	gcatgttcat	catcagtaac	3000
ccgtatcgtg	agcatcctct	ctcgtttcat	gggtatcatt	acccccatga	acagaaatcc	3060
cccttacacg	gaggcatcag	tgaccaaaaa	ggaaaaaac	gcccttaaca	tgcccgcgtt	3120
tatcagaagc	cagacattaa	cgcttctgga	gaaactcaac	gagctggacg	cggtatgaac	3180
ggcagacatc	tgtgaatcgc	ttcacgacca	cgctgatgag	ctttaccgca	gctgcctcgc	3240
gcgtttcggt	gatgacgggt	aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	3300
ttgtctgtaa	gcggatgccg	ggagcagaca	agcccgtcag	ggcgcgctcag	cggtgttggt	3360
cgggtgtcgg	ggcgagacca	tgaccagtc	acgtagcgat	agcggagtgt	atactggjct	3420
aactatcgcg	catcagagca	gattgtactg	agagtgcacc	atatatgcgg	tgtgaaatac	3480
cgcacagatg	cgtaaggaga	aaataccgca	tcaggcgctc	ttccgcttcc	tcgctcactg	3540
actcgctcgg	ctcggtcggt	cggctgcggc	gagcgggtatc	agctcactca	aaggcggtaa	3600
tacgggttatc	cacagaatca	ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	3660
aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	3720
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagagggtg	gcgaaaccgc	acaggactat	3780
aaagatacca	ggcgtttccc	cttggaagct	ccctcgctgc	ctctcctggt	ccgaccctgc	3840
cgcttacccg	atacctgtcc	gcctttctcc	cctcggggaag	cgtggcgctt	tctcatagct	3900
cacgctgtag	gtatctcagt	tcgggtgtag	tcgttcgctc	caagctgggc	tgtgtgcacg	3960
aaccccccg	tcagcccagc	cgctgcgcct	tatccggtaa	ctatcgtctt	gagtccaacc	4020
cggtaagaca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	4080
ggatgtagg	cggtgctaca	gagttcttga	agtgggtggc	taactacggc	tacactagaa	4140
ggacagtatt	tggtatctgc	gctctgtcga	agccagttac	cttcggaaaa	agagttggta	4200
gctcttgatc	cggcaaaaca	accaccgctg	gtagcgggtg	tttttttggt	tgcaagcagc	4260
agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	gatcttttct	acgggggtctg	4320
acgctcagtg	gaacgaaaac	tcacgttaag	ggattttggt	catgagatta	tcaaaaagga	4380
tcttcaccta	gatcctttta	aattaaaaat	gaagttttaa	atcaatctaa	agtatatatg	4440
agtaaaactg	gtctgacagt	taccaatgct	taatcagtga	ggcacctatc	tcagcgatct	4500
gtctatttcg	ttcatccata	gttgccctgac	tcctcgctcg	gtagataact	acgatacggg	4560
agggcttacc	atctggcccc	agtgctgcaa	tgataccgcg	agaccacgc	tcaccggctc	4620
cagatttatc	agcaataaac	cagccagccg	gaagggccga	gcgcagaagt	ggtcctgcaa	4680
ctttatccgc	ctccatccag	tctattaatt	gttgccggga	agctagagta	agtagttcgc	4740
cagttaatag	tttgcgcaac	gttggtgcca	ttgctgcagg	catcggtgtg	tcacgctcgt	4800
cgtttggtat	ggcttcattc	agctccgggt	cccaacgatc	aaggcgagtt	acatgatccc	4860
ccatgtgtg	caaaaaagcg	gttagctctc	tcggctctcc	gatcggtgtc	agaagtaagt	4920
tggcgcaggt	gttatcactc	atggttatgg	cagcactgca	taattctctt	actgtcatgc	4980
catccgtaag	atgcttttct	gtgactgggt	agtactcaac	caagtcattc	tgagaatagt	5040
gtatgcggcg	accgagttgc	tcttgcccg	cgtcaatcag	ggataatacc	gcgccacata	5100
gcagaacttt	aaaagtgtc	atcattggaa	aacgttcttc	ggggcgaaaa	ctctcaagga	5160

```

tettaccgct gttgagatcc agttcgatgt aaccactcg tgcacccaac tgatcttcag 5220
catcttttac tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa aatgccgcaa 5280
aaaagggaat aaggcgacac cggaaatgtt gaatactcat actcttcctt tttcaatatt 5340
attgaagcat ttatcagggt tattgtctca tgagcggata catatttgaa tgtattttaga 5400
aaaataaaca aatagggtt ccgcgcacat tccccgaaa agtgccacct gaaattgtaa 5460
acgttaatat tttgttaaaa ttccggttaa atttttgtta aatcagctca ttttttaacc 5520
aataggccga aatcggaaca atcccttata aatcaaaaga atagaccgag atagggttga 5580
gtgttgttcc agtttggaa aagagtccac tattaagaa cgtggactcc aacgtcaaag 5640
ggcgaaaaac cgtctatcag ggcgatggcc cactacgtga accatcacc taatcaagtt 5700
tttgggggtc gaggtgccgt aaagcactaa atcggaaccc taaaggagc ccccgattta 5760
gagcttgacg gggaaagccg gcgaacgtgg cgagaaagga agggaaagaa gcgaaaggag 5820
egggcgctag ggcgctggca agtgtagcgg tcacgctgcg cgtaaccacc acaccgccc 5880
cgcttaatgc gccgctacag ggcgctgcc attcgcca 5918

```

<410> 43

<411> 5891

<412> ADN

<413> Séquence artificielle

<420>

<423> Description de la séquence artificielle: vecteur
d'expression pET32a-dp-Pt (TME2)

<400> 43

```

atccggatat agttctctct ttcagcaaaa aaccctcaa gaccggttta gagggcccaa 60
ggggttatgc tagttattgc tcagcgggtg cagcagccaa ctacagcttc tttcgggctt 120
tgtagcagc cggatctcag tgggtggtgg ggtggtgctc gagtgccggc gcaagcttgt 180
cgacggagct cgaattccta agcttcagcc tgagagatca gcagcatcat ccacaggcaa 240
gagcaaacac gagcgtcagc cagcagcagg aacagcagaa caacgtattc cgggtcggat 300
ccacgcggaa ccagaccacc accaccaca gacgatcag atccagaacc agaaccggcc 360
aggttagcgt cgaggaactc tttcaactga cctttagaca gtgcacccac tttggttgcc 420
gccacttcac cgtttttgaa cagcagcaga gtcgggatac cagcgatgcc atatttcggc 480
gcagtgccag ggttttgatc gatgttcagt tttgcaacgg tcagtttgcc ctgatattcg 540
tcagcgattt catccagaat cggggcgatc attttgacg gaccgcacca ctctgccag 600
aaatcgacga ggatcgcccc gtcgctttg agtacatccg tgtcaaaact gtcgtcagtc 660
aggtgaataa tttttcgct catatgtata tctcttctt aaagttaaac aaaattattt 720
ctagagggga attgttatcc gctcacaatt cccctatagt gagtcgtatt aatttcgagg 780
gatecagatc gatctcgatc ctctacgccg gacgcacgt ggccggcatc accggcgcca 840
caggtgcggg tgctggcgcc tatatcgccg acatcaccga tggggaagat cgggctcgcc 900
acttcgggct catgagcgct tgtttcggcg tgggtatggt ggcaggcccc gtggccgggg 960
gactgttggg cgccatctcc ttgcatgcac cattccttgc ggcggcgggt ctcaacggcc 1020
tcaacctact actgggctgc ttcctaattg aggagtcgca taaggagag cgtcgagatc 1080
ccggacacca tcgaatggcg caaaaccttt cgcggtatgg catgatagcg cccggaagag 1140
agtcaattca ggggtgtgaa tgtgaaacca gtaacgttat acgatgtcgc agagtatgcc 1200
gggtgtctct atcagaccgt tccccgcgtg gtgaaccagg ccagccacgt ttctgcgaaa 1260
acgcgggaaa aagtggaaagc ggcgatggcg gagctgaatt acattcccaa ccgcgtggca 1320
caacaactgg cgggcaaaaca gtcgttgctg attggcgttg ccacctccag tctggccctg 1380
cacgcgcgt cgcaaatgt cgcgcgat aaatctcgcg ccgatcaact gggtgccagc 1440
gtggtggtgt cgatggtaga accgaagcgt gtcgaagcct gtaaagcggc ggtgcacaat 1500
cttctcgcg aacgcgtcag tgggctgac attaaactac cgctggatga ccaggatgcc 1560
attgctgtgg aagctgcctg cactaatgtt ccggcggtat ttcttgatgt ctctgaccag 1620
acacccatca acagtattat tttctcccat gaagacggta cgcgactggg cgtggagcat 1680
ctggctcgat tgggtcacca gcaaatcgcg ctgttagcgg gccatttaag ttctgtctcg 1740
gcgcgtctgc gtctggctgg ctggcataaa tatctcact gcaatcaaat tcagccgata 1800
cggaagacggg aaggcgactg gagtgcctat tccggttttc aacaaaccat gcaaatgctg 1860
aatagggca tcgttccac tgcgatgctg gttgccaacg atcagatggc gctggcgca 1920
atgcgcgcca ttaccgagtc cgggctgcgc gttggtgcgg acatctcggg agtgggatac 1980
gacgataccg aagacagctc atgttatatc ccgcggttaa ccaccatcaa acaggatttt 2040
cgctgtctgg ggcaaacag cgtggaccgc ttgctgcaac tctctcaggg ccaggcgggtg 2100
aagggaatc agctgttgcc cgtctcactg gtgaaaagaa aaaccaccct ggcgccaat 2160

```


acgcaaaccg	cctctccccg	cgcggtggcc	gattcattaa	tgcagctggc	acgacaggtt	2220
tcccgaactg	aaagcgggca	gtgagcgcaa	cgcaattaat	gtaagttagc	tcactcatta	2280
ggcaccggga	tctcgaccga	tgcccttgag	agccttcaac	ccagtcagct	ccttccggtg	2340
ggcgcggggc	atgactatcg	tcgccgcact	tatgactgtc	ttctttatca	tgcaactcgt	2400
aggacaggtg	ccggcagcgc	tctgggtcat	tttcggcgag	gaccgctttc	gctggagcgc	2460
gacgatgac	ggcctgtcgc	ttgcggtatt	cggaatcttg	cacgacctcg	ctcaagcctt	2520
cgtcactgg	cccgccacca	aacgtttcgg	cgagaagcag	gccattatcg	ccggcatggc	2580
ggccccacgg	gtgcgcatga	tctgtctcct	gtcgttgagg	acccggctag	gctggcgggg	2640
ttgccttact	ggtttagcaga	atgaatcacc	gatacgcgag	cgaacgtgaa	gcgactgctg	2700
ctgcaaaacg	tctgcgacct	gagcaacaac	atgaatgggc	ttcgggtttc	gtgtttcgta	2760
aagtctggaa	acgcggaagt	cagcgccctg	caccattatg	ttccggatct	gcacgcgagg	2820
atgctgctgg	ctaccctgtg	gaacacctac	atctgtatta	acgaagcgt	ggcattgacc	2880
ctgagtgatt	tttctctggt	cccgccgcat	ccataccgcc	agttgtttac	cctcacaacg	2940
ttccagtaac	cgggcatggt	catcatcagt	aaccgcgtatc	gtgagcatcc	tctctcgttt	3000
catcgggtatc	attacccccca	tgaacagaaa	tcccccttac	acggaggcat	cagtgaccaa	3060
acaggaaaaa	accgccctta	acatggcccg	ctttatcaga	agccagacat	taacgcttct	3120
ggagaaactc	aacgagctgg	acgcggatga	acaggcagac	atctgtgaat	cgcttcacga	3180
ccacgctgat	gagctttacc	gcagctgcct	cgcgcggttc	ggtgatgacg	gtgaaaacct	3240
ctgacacatg	cagctcccgg	agacggtcac	agcttgtctg	taagcggatg	ccgggagcag	3300
acaagcccg	cagggcgcg	cagcggtgtg	tggcggtgtg	cgggcgcgag	ccatgaccca	3360
gtcacgtagc	gatagcggag	tgtatactgg	cttaactatg	cggcacacga	gcagattgta	3420
ctgagagtg	accatatatg	cggtgtgaaa	taccgcacag	atgcgtaagg	agaaaatacc	3480
gcatcaggcg	ctcttccgct	tcctcgctca	ctgactcgct	gcgctcggtc	gttcggctgc	3540
ggcgagcggt	atcagctcac	tcaaaggcgg	taatacggtt	atccacagaa	tcaggggata	3600
acgcagga	gaacctgtga	gcaaaaggcc	cgcaaaagc	caggaaccgt	aaaaaggccg	3660
cgttgctggc	gtttttccat	aggctccgcc	cccctgacga	gcatcacaaa	aatcgacgct	3720
caagtcagag	gtggcgaaac	ccgacaggac	tataaagata	ccaggcggtt	ccccctggaa	3780
gctccctcgt	gcgctctcct	gttccgaccc	tgccgcttac	cggatacctg	tccgcctttc	3840
tcccttcggg	aagcgtggcg	ctttctcata	gctcacgctg	taggtatctc	agttcggtgt	3900
aggctggtcg	ctccaagctg	ggctgtgtgc	acgaaccccc	cgttcagccc	gaccgctgcg	3960
ccttatccgg	cttgagtcca	acccggtaa	acacgactta	tcgccactgg	tcgccactgg	4020
cagcagccac	tggtaacagg	attagcagag	cgaggtatgt	aggcggtgct	acagagttct	4080
tgaagtgggt	gcctaactac	ggctacacta	gaaggacagt	atttggtatc	tgcgctctgc	4140
tgaagccagt	taccttcgga	aaaagagttg	gtagctcttg	atccggcaaa	caaaccaccg	4200
ctggtagcgg	tggttttttt	gtttgcaagc	agcagattac	gcgcagaaaa	aaaggatctc	4260
aagaagatcc	tttgatcttt	tctacggggg	ctgacgctca	gtggaacgaa	aactcacggt	4320
aagggttttt	ggctatgaga	ttatcaaaaa	ggatcttcac	ctagatcctt	ttaaattaaa	4380
aatgaagttt	taaatcaatc	taaagtatat	atgagtaaac	ttggtctgac	agttaccaat	4440
gcttaatcag	tgaggcacct	atctcagcga	tctgtctatt	tcgttcatcc	atagttgcct	4500
gactccccgt	cgtgtagata	actacgatac	gggagggctt	accatctggc	cccagtgtctg	4560
caatgatacc	gcgagaccca	cgctcacccg	ctccagattt	atcagcaata	aaccagccag	4620
ccggaagggc	cgagcgcaga	agtggctcctg	caactttatc	cgctccatc	cagtctatta	4680
attggtgcg	ggaagctaga	gtaagtagtt	cgccagttaa	tagtttgccg	aacgttggtg	4740
ccattgctgc	aggcatcggtg	gtgtcacgct	cgctggttgg	tatggcttca	ttcagctccg	4800
gttcccaacg	atcaaggcga	gttacatgat	cccccatggt	gtgcaaaaaa	gcgggttagct	4860
ccttcgggtc	tccgatcggt	gtcagaagta	agttggccgc	agtgttatca	ctcatgggtta	4920
tggcagcact	gcataattct	cttactgtca	tgccatccgt	aagatgcttt	tctgtgactg	4980
gtgagtactc	aaccaagtca	ttctgagaat	agtgtatgcg	gcgaccgagt	tgctcttgcc	5040
cggcgctca	acgggataat	accgcgccac	atagcagaac	tttaaaagt	ctcatcattg	5100
gaaaacgttc	ttcggggcga	aaactctcaa	ggatcttacc	gctggttgaga	tccagttcga	5160
tgtaaccac	tcgtgcaccc	aactgatctt	cagcatcttt	tactttcacc	agcgtttctg	5220
ggtgagcaaa	aacaggaagg	caaaatgccg	caaaaaagg	aataaggcg	acacggaaat	5280
gttgaatact	catactcttc	ctttttcaat	attattgaag	catttatcag	ggttattgtc	5340
tcatgagcgg	atacatatct	gaatgtatct	agaaaaataa	acaaatagg	gttccgcgca	5400
caatttcccc	aaaagtgcga	cctgaaattg	taaaacgtta	tattttgtta	aaattcgcgt	5460
taaaattttg	ttaaatcagc	tcaattttta	accaataggc	cgaaatcggc	aaaatccctt	5520
ataaatcaaa	agaatagacc	gagatagggt	tgagtgttgt	tccagtttgg	aacaagagtc	5580
cactattaaa	gaacgtggac	tccaacgtca	aagggcgaaa	aaccgtctat	cagggcgatg	5640
gcccactacg	tgaaccatca	ccctaataca	gttttttggg	gtcgaggtgc	cgtaaaagcac	5700
taaatcgga	ccctaagggg	agccccgat	ttagagcttg	acggggaaa	ccggcgaaacg	5760
tggcgagaaa	ggaagggaag	aaagcgaaag	gagcgggcgc	tagggcgctg	gcaagtgtag	5820

cggtcacgct gcgcgtaacc accacacccg ccgcgcttaa tgcgcgcta cagggcgcggt 5880
cccattcgcc a 5891

<210> 44

<211> 2617

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: vecteur
d'expression pT7-7-dp-Pt (TME1)

<400> 44

```

aattctcatg tttgacagct tatcatcgat gataagcttg ggctgcaggt cgactctaga 60
ggatccccgg gcgcgaattc ctaagcgtca acaccagcga acagcagcag aacaaccaga 120
actttagccc agttaccaac catagagaag taagcgatac cagccagaac accccagtga 180
gcaccagcga tcgggtccat atgtatatct ccttcttaaa gttaaacaaa attatttcta 240
gagggaacac gttgtggtct ccctatagtg agtcgtatta atttcgaagt ctatcagaag 300
ttcgaatcgc tgggctccgc gcgtttcggt gatgacggtg aaaacctctg acacatgcag 360
ctcccggaga cggtcacagc ttgtctgtaa gcggatgccg ggagcagaca agcccgtcag 420
ggcgcgtcag cgggtgttgg cgggtgtcgg ggccgagcca tgaccagtc acgtagcgat 480
agcggagtgt atatactggc ttaactatgc ggcatcagag cagattgtac tgagagtgc 540
ccataggaag atcttccgga agatcttccct atgcggtgtg aaataccgca cagatgcgta 600
aggagaaaat accgcacagc gcgctcttcc gcttccctgc tcaactgact gctgcgctcg 660
gtcgttcggc tgcggcgagc tcgctccaag cactcaaaag cggtaatcag gttatccaca 720
gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac 780
cgtaaaaagg ccggtgtgtt ggcgtttttc cataggctcc gccccctga cgagcatcac 840
aaaaatcgac gctcaagtca gaggtggcga aaccgcagag gactataaag ataccaggcg 900
tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgccgct taccggatac 960
ctgtccgctt ttctcccttc gggaagcgtg gcgcttctc aatgctcacg ctgtaggat 1020
ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc cccggttcag 1080
cccagccgct gcgccttata cggttaactat cgtcttgagt ccaaccgggt aagacacgac 1140
ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta tgtaggcgg 1200
gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttgg 1260
atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccgg 1320
aaacaaacca ccgctggtag cgggtggtttt tttgtttgca agcagcagat tacgcgcaga 1380
aaaaaaggat ctcaagaaga tcctttgatc tttctacgg ggtctgacgc tcagtggaa 1440
gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt cacctagac 1500
cttttaattc ttgaagacga aagggcctcg tgatacgctt atttttatag gttaatgtc 1560
tgataataat ggtttcttag acgtcaggtg gcacttttcg gggaaatgtg cgcggaacc 1620
ctatttgttt atttttctaa atacattcaa atatgtatcc gctcatgaga caataacc 1680
gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat ttccgtgtc 1740
cccttattcc cttttttgcg gcattttgct ttctgtttt tgctcaccca gaaacgctg 1800
tgaaagtaaa agatgctgaa gatcagttgg gtgcacagat ggggttacatc gaactggat 1860
tcaacagcgg taagatcctt gagagttttc gccccgaaga acgttttcca atgatgagc 1920
cttttaaaag tctgctatgt ggccggttat tatcccggtg tgacgccggg caagagcaac 1980
tcggtcgccg catacactat tctcagaatg acttggttga gtactacca gtcacagaaa 2040
agcatcttac ggatggcatg acagtaagag aattatgcag tgctgccata accatgagt 2100
ataacactgc ggccaactta cttctgacaa cgatcggagg accgaaggag ctaaccgctt 2160
ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg gagctgaat 2220
aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca acaacgttg 2280
gcaaaactatt aactggcgaa ctacttactc tagcttcccg gcaacaatta atagactgg 2340
tgaggcgga taaagtgtgca ggaccacttc tgcgctcggc ccttccggct ggctggtt 2400
ttgctgataa atctggagcc ggtgagcgtg ggtctcggcg tatcattgca gcaactatg 2460
cagatggtaa gccctcccgt atcgtagtta tctacacgac ggggagtcag gcaactatg 2520
atgaacgaaa tagacagatc gctgagatga gtgcctcact gattaagcat tggttaact 2580
cagaccaagt ttactcatat atactttaga ttgattt 2617

```

<210> 45

<211> 2599

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: vecteur d'expression pT7-7-dp-Pt (TME2)

<400> 45

```

aattctcatg tttgacagct tatcatcgat gataagcttg ggctgcaggt cgactctaga 60
ggatccccgg gcgcgaattc ctaagcttca gcctgagaga tcagcagcat catccacagg 120
caagagcaaa cacgagcgctc agccagcagc aggaacagca gaacaacgta ttccgggtcc 180
atatgtatat ctcccttctta aagttaaaca aaattatttc tagagggaaa ccgttggtgt 240
ctccctatag tgagtcgtat taatttcgaa gtctatcaga agttcgaatc gctgggcctc 300
gcgcgtttcg gtgatgacgg tgaaaacctc tgacacatgc agctcccgga gacgggtcac 360
gcttgctctgt aagcgggatgc cgggagcaga caagcccgtc agggcgcgctc agcgggtgtt 420
ggcgggtgtc gggggcgagc catgaccagc tcacgtagcg atagcggagt gtatatactg 480
gcttaactat gcggcatcag agcagattgt actgagagt caccatagga agatcttccg 540
gaagatcttc ctatgcgggtg tgaaataccg cacagatgcg taaggagaaa ataccgcac 600
agcgctcttc ccgcttcttc gctcactgac tcgctgcgct cggctcgttc gctgcggcga 660
gcgggtatcag ctcaactcaa ggcggtaata cggttatcca cagaatcagg ggataacgca 720
ggaaagaaca tgtgagcaaa aggcagcaca aaggccagga accgtaaaaa ggccgcgttg 780
ctggcggtttt tccataggct ccgccccctc gacgagcatc acaaaaatcg acgctcaagt 840
cagaggtggc gaaacccgac aggaactataa agataccagg cgtttccccc tgggaagctcc 900
ctcgtgcgct ctccgtgttc gaccctgccc cttaccggat acctgtccgc ctttctccct 960
tcgggaagcg tggcgctttc tcaatgctca cgctgtaggt atctcagttc ggtgtaggtc 1020
gttcgctcca agctgggctg tgtgcacgaa cccccgctc agcccgaccg ctgcgcctta 1080
tccggttaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc actggcagca 1140
gccactggtt acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag 1200
tgggtggccta actacggcta cactagaagg acagtatttg gtatctgcgc tctgctgaag 1260
ccagttacct tcggaaaaag agttggtagc tcttgatccc gcaaaaaaac caccgctggg 1320
agcggtggtt tttttgtttg caagcagcag attacgcgca gaaaaaagg atctcaagaa 1380
gatcctttga tcttttctac ggggtctgac gctcagtgga acgaaaactc acgttaaggg 1440
atthtgggtca tgagattatc aaaaaggatc ttcacctaga tctttttaat tcttgaagac 1500
gaaagggcct cgtgatacgc ctatttttat aggttaatgt catgataata atggtttctt 1560
agacgtcagg tggcactttt cggggaaatg tgcgcggaac ccctatttgt ttatttttct 1620
aaatacatc aaatatgtat ccgctcatga gacaataacc ctgataaatg cttcaataat 1680
attgaaaaag gaagagtatg agtattcaac atttccgtgt cgcccttatt cccttttttg 1740
cggcattttg ccttccgtgt tttgctcacc cagaaacgct ggtgaaagta aaagatgctg 1800
aagatcagtt ggggtgcacga gtgggttaca tcgaactgga tctcaacagc ggtaagatcc 1860
ttgagagttt tcgccccgaa gaacgttttc caatgatgag cactttttaa gttctgctat 1920
gtggcgcggt attatcccggt gttgacgcgc ggcaagagca actcggtcgc cgcatacact 1980
attctcagaa tgacttggtt gagtaactcac cagtcacaga aaagcatctt acggatggca 2040
tgacagtaag agaattatgc agtgcgtgca taacctagag tgataaacact gcggccaact 2100
tacttctgac aacgatcgga ggaccgaagg agctaaccgc ttttttgcac aacatggggg 2160
atcatgtaac tcgccttgat cggtgggaac cggagctgaa tgaagccata ccaaacgacg 2220
agcgtgacac cacgatgcct gtagcaatgg caacaacgtt gcgcaacta ttaactggcg 2280
aactacttac tctagcttcc cggcaacaat taatagactg gatggaggcg gataaagttg 2340
caggaccact tctgcgctcg gcccttccgg ctggctggtt tattgctgat aaatctggag 2400
ccggtgagcg tgggtctcgc ggtatcattg cagcactggg gccagatggt aagccctccc 2460
gtatcgtagt tatctacacg acggggagtc aggcaactat ggatgaacga aatagacaga 2520
tcgctgagat aggtgcctca ctgattaagc attggttaact gtcagacca gtttactcat 2580
atatacttta gattgattt

```

<210> 46

<211> 271

<212> PRT

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: protéine
de fusion GST-DP-TME1

<400> 46

```

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro
 1           5           10           15

Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu
          20           25           30

Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu
          35           40           45

Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
          50           55           60

Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn
          65           70           75           80

Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu
          85           90           95

Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser
          100          105          110

Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu
          115          120          125

Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn
          130          135          140

Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp
          145          150          155          160

Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu
          165          170          175

Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr
          180          185          190

Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala
          195          200          205

Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Ser Gly Gly
          210          215          220

Gly Gly Gly Leu Val Pro Arg Gly Ser Asp Pro Ile Ala Gly Ala His
          225          230          235          240

Trp Gly Val Leu Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp
          245          250          255

Ala Lys Val Leu Val Val Leu Leu Leu Phe Ala Gly Val Asp Ala
          260          265          270

```

<210> 47

<211> 265

<212> PRT

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: protéine
de fusion GST-DP-TME2

<400> 47

```

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro
 1           5           10           15

Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu
      20           25           30

Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu
      35           40           45

Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
      50           55           60

Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn
      65           70           75           80

Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu
      85           90           95

Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser
      100          105          110

Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu
      115          120          125

Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn
      130          135          140

Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp
      145          150          155          160

Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu
      165          170          175

Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr
      180          185          190

Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala
      195          200          205

Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Ser Gly Gly
      210          215          220

Gly Gly Gly Leu Val Pro Arg Gly Ser Asp Pro Glu Tyr Val Val Leu
      225          230          235          240

Leu Phe Leu Leu Leu Ala Asp Ala Arg Val Cys Ser Cys Leu Trp Met
      245          250          255

Met Leu Leu Ile Ser Gln Ala Glu Ala
      260          265

```

<210> 48

<211> 170

<212> PRT

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: protéine
de fusion TrX-DP-TME1

<400> 48

Met Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp
1 5 10 15Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp
20 25 30Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp
35 40 45Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn
50 55 60Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu
65 70 75 80Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser
85 90 95Lys Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala Gly Ser Gly
100 105 110Ser Gly Ser Pro Lys Ser Asp Leu Ser Gly Gly Gly Gly Leu Val
115 120 125Pro Arg Gly Ser Asp Pro Ile Ala Gly Ala His Trp Gly Val Leu Ala
130 135 140Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val Leu Val
145 150 155 160Val Leu Leu Leu Phe Ala Gly Val Asp Ala
165 170

<210> 49

<211> 161

<212> PRT

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: protéine
de fusion TrX-DP-TME2

<400> 49

Met Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp
1 5 10 15Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp
20 25 30Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp
35 40 45Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn
50 55 60

Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu
65 70 75 80
Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser
85 90 95
Lys Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala Gly Ser Gly
100 105 110
Ser Gly Ser Asp Leu Ser Gly Gly Gly Gly Leu Val Pro Arg Gly
115 120 125
Ser Asp Pro Glu Tyr Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala
130 135 140
Arg Val Cys Ser Cys Leu Trp Met Met Leu Leu Ile Ser Gln Ala Glu
145 150 155 160
Ala

<210> 50

<211> 39

<212> PRT

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: protéine
de fusion M-DP-TME1

<400> 50

Met Asp Pro Ile Ala Gly Ala His Trp Gly Val Leu Ala Gly Ile Ala
1 5 10 15

Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val Leu Val Val Leu Leu
20 25 30

Leu Phe Ala Gly Val Asp Ala
35

<210> 51

<211> 33

<212> PRT

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: protéine
de fusion M-DP-TME2

<400> 51

Met Asp Pro Glu Tyr Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala
1 5 10 15

Arg Val Cys Ser Cys Leu Trp Met Met Leu Leu Ile Ser Gln Ala Glu
20 25 30

Ala

<210> 52
 <211> 239
 <212> PRT
 <213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:
 glutathion transférase (GST)

<400> 52

```

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro
 1           5           10           15

Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu
          20           25           30

Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu
          35           40           45

Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
          50           55           60

Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn
          65           70           75           80

Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu
          85           90           95

Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser
          100          105          110

Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu
          115          120          125

Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn
          130          135          140

Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp
          145          150          155          160

Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu
          165          170          175

Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr
          180          185          190

Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala
          195          200          205

Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Ser Gly Gly
          210          215          220

Gly Gly Gly Leu Val Pro Arg Gly Ser Pro Gly Ile His Arg Asp
          225          230          235

```

<210> 53
 <211> 170
 <212> PRT

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:
thiorédoxine (TrX)

<400> 53

```

Met Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp
 1           5           10           15

Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp
          20           25           30

Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp
          35           40           45

Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn
 50           55           60

Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu
 65           70           75           80

Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser
          85           90           95

Lys Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala Gly Ser Gly
          100          105          110

Ser Gly Ser Pro Lys Ser Asp Leu Ser Gly Gly Gly Gly Gly Leu Val
          115          120          125

Pro Arg Gly Ser Asp Pro Ile Ala Gly Ala His Trp Gly Val Leu Ala
          130          135          140

Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val Leu Val
          145          150          155          160

Val Leu Leu Leu Phe Ala Gly Val Asp Ala
          165          170

```